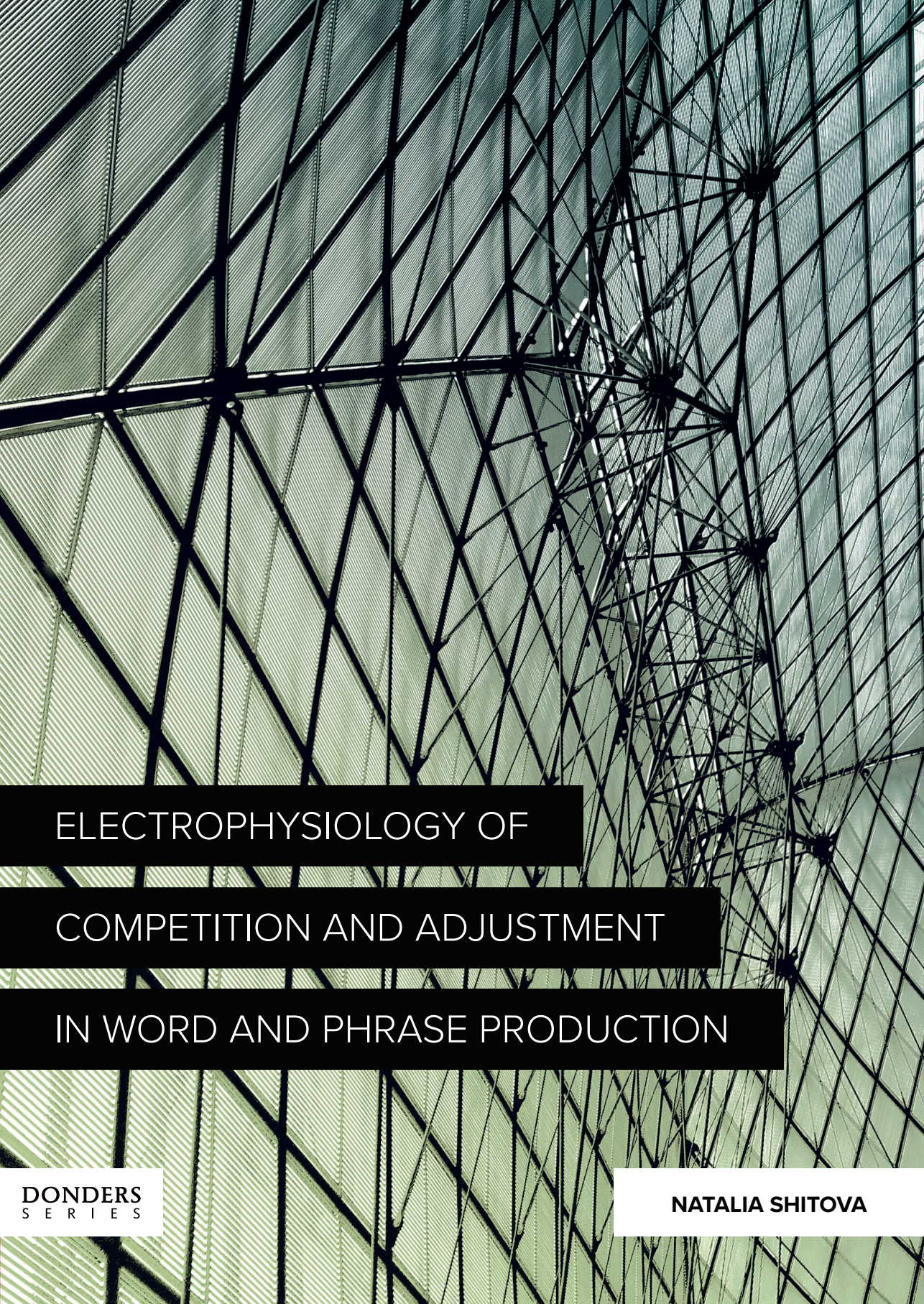

ELECTROPHYSIOLOGY OF COMPETITION AND ADJUSTMENT IN WORD AND PHRASE PRODUCTION

DONDERS SERIES

NATALIA SHITOVA

ELECTROPHYSIOLOGY OF COMPETITION AND ADJUSTMENT IN WORD AND PHRASE PRODUCTION

NATALIA SHITOVA

Cover artist:	Jan Maas	
Cover photograph:	Natalia Shitova	
Layout & cover design:	Design Your Thesis	www.designyourthesis.com
Print:	Ridderprint	www.ridderprint.nl
ISBN:	978-94-6284-139-0	

ELECTROPHYSIOLOGY OF COMPETITION AND ADJUSTMENT IN WORD AND PHRASE PRODUCTION

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J. H. M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 5 juni 2018
om 12:30 uur precies

door

Natalia Shitova
geboren op 12 oktober 1988
te Arzamas-16, Sovjet Unie

Promotoren:

Prof. dr. A. P. A. Roelofs
Prof. dr. H. J. Schriefers

Copromotoren:

Dr. J. M. Schoffelen
Dr. M. C. M. Bastiaansen

Manuscriptcommissie:

Prof. dr. A. S. Meyer
Prof. dr. P. Indefrey (Universität Düsseldorf)
Dr. K. M. Lemhöfer

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Producing language relies on a coordinated sequence of linguistic, motor, and general attentional processes. These processes can be probed selectively in an experimental setup, as well as affected in speech and language disorders. One of the essential linguistic processes associated with language production is lexical selection, which involves selection of relevant words from the lexicon to be used in speaking. A prominent theory of language production (Levelt, Roelofs, & Meyer, 1999) assumes that lexical selection is a competitive process, i.e., every produced word is a result of comparison of activation levels of multiple alternatives, which "compete" to be selected. In this dissertation, I provide new evidence supporting this view. Furthermore, I demonstrate that speakers can strategically adjust parameters of this competitive selection mechanism depending on their expectations of the task that they are performing at the moment. In the Introduction to this dissertation I first explain the competition account in spoken word production and present experimental evidence that supports this approach or challenges it. Then I outline situations in which based on available information a speaker can adjust control in lexical selection. I further elaborate on adjustments that a speaker employs in a switching task.

1.1. COMPETITION IN WORD SELECTION

A theory of language production developed by Levelt, Roelofs, and Meyer (1999; Roelofs, 1992, 2003) posits that concept-based production of a word takes three main stages: conceptualization, word planning, and articulation. During the first stage an existing concept needs to be identified in long-term memory. During the word planning stage a lemma is selected from the mental lexicon (*lexical selection*), its morphological and phonological properties are retrieved (*grammatical* and *phonological encoding*) and the phonological information is translated into motor commands (*phonetic encoding*) that will be executed during the last stage of articulation. These processes act on a network of nodes of multiple levels. Within the network, activation spreads vertically (from high-level conceptual nodes through lexical nodes down to morphemes, phonemes, and finally syllable motor programs), as well as horizontally (e.g., from a concept to semantically related concepts). In absence of such horizontal connections word production would be streamlined all the way from concepts to articulatory programs. However, the horizontal spreading activation allows other related concepts to get activated and initiate propagation of activation downstream to lexical, grammatical, phonological, and articulational nodes, associated with these concepts. Using an example from the original article (Levelt et al., 1999), activation of a lexical concept BED will activate not only its subordinate nodes, i.e., a lemma *bed* and further representations at morphological, phonological and phonetical levels, but also a related concept CHAIR, which will lead to activation of a lemma *chair* and its further representations. Therefore, in order to be selected the lemma *bed* will have to compete with other lemmas (e.g., *chair*, *desk*, *closet*),

activated by lexical concepts related to the originally identified concept "BED". The notion of lexical selection by competition will be further discussed in the context of specific language production tasks.

The simplest case of language production – production of single words – is often operationalized in a lab setup through the picture naming task. Drawings of objects are presented to a participant one by one and requested to be named as quickly and as accurately as possible. This task is widely used in clinical applications (e.g., Hashimoto & Thompson, 2010; Seiger-Gardner & Schwartz, 2008; Thompson, Cho, Price, Wieneke, Bonakdarpour, & Rogalski, 2012), as well as in studies of language acquisition and learning (e.g., Carpenter & Olson, 2012). It is generally assumed that perception of an image directly activates the lexical concept associated with the features of the picture. The timing of the production processes for a picture naming task with an average reaction time (RT) of 600 was identified by Indefrey and Levelt (2004; Indefrey, 2011). The perceptual and conceptual encoding stage takes the first 200 ms, followed by lexical selection that takes the next 75 ms. The remainder of the word planning stage takes different amount of time depending on the length of the word to be produced in syllables: A two-syllable word, for instance, would take approximately 325 ms to complete word planning. In case the RT is longer than 600 ms the timing estimates of the stages need to be rescaled.

The picture naming task can be modified in order to target specific research questions in language production. For example, the picture-word interference (PWI; Glaser & Düngelhoff, 1984) paradigm employs naming of pictures that have superimposed distractor words on them. It is assumed that while a picture activates relevant concepts in the long-term memory and triggers activation of the relevant subordinate nodes (i.e., lemmas, morphemes, phonemes, etc.), a distractor word activates the corresponding lemma directly. Manipulating the relation between the concept illustrated by a picture and the distractor word allows to assess competition in lexical selection.

A standard RT pattern usually observed in a PWI task is as follows. If the distractor word is *congruent* with the object in the picture (e.g., the word *dog* superimposed on a picture of a dog) the naming is faster and more accurate than if the distractor word is *incongruent* with the object (e.g., the word *cat* superimposed on a picture of a dog). Moreover, the congruent trials take less time and the incongruent trials take more time relatively to stimuli in which the distractor word is unrelated to the picture name (e.g., the word *desk* superimposed on a picture of a dog) or in which the distractor is a meaningless sequence of letters (e.g., a sequence XXX superimposed on a picture of a dog). The difference in RT associated with congruent vs. incongruent trials in PWI is called the Stroop-like effect. Indeed, the PWI task appears very similar to the classical color-word Stroop paradigm, in which a speaker

names ink color of color names, while the relation between ink color and the color name is manipulated. If the color name is congruent with the ink color (e.g., the word *red* written in red ink) the naming is faster and more accurate than if the color name is incongruent with the ink color (e.g., the word *blue* written in red ink).

The theoretical framework (Levelt et al., 1999; Roelofs, 2014) explains these behavioral patterns through competitive lexical selection. On a congruent PWI trial both picture and distractor activate the same lemma most strongly (e.g., a lemma 'dog' in the previous example). Due to horizontal spreading activation, lemmas corresponding to related lexical concepts (e.g., 'cat', 'bird', 'fish') will also be somewhat activated, but this activation will be substantially smaller than the joint activation of the target lemma, therefore, the target lemma will easily win the competition against other alternatives. On an incongruent PWI trial the lexical concept activated by the picture will again activate the target lemma, as well as a network of semantically-related concepts (via horizontal connections) and, consequently, lemmas associated with these semantically-related concepts. The distractor word will directly activate its lemma (which is also one of lemmas that received activation from the picture through activation of a semantically-related concept). The distractor lemma will furthermore activate its concept via feedback connections, which in its turn will add to the activation of the concept linked to the picture. As a result of such a complex pattern of activations, the distractor lemma receives a level of activation that is comparable with the level of activation received by the target lemma. Therefore, performance on incongruent trials remarkably drops compared to congruent trials: It takes longer for a participant to give a response and there is a higher chance for an error. If a distractor is unrelated to the picture, the distractor lemma receives activation only from the distractor itself, but not from the target concept via horizontal connections to semantically-related concepts, as in the case of incongruent trials. Therefore, although the distractor lemma enters competition with the target lemma, its activation level is rather small and performance of such trials is better than on incongruent trials. Lastly, if a distractor is a non-word sequence of letters, no lemma gets activated by the distractor and, similarly to congruent trials, no competition takes place. However, in case of a neutral distractor, the distractor does not help activation of the target lemma, therefore performance is worse than performance on congruent trials.

The competitive lexical selection approach appears to be supported by experimental evidence, as well as it showed its utility in computer modeling (Roelofs, 1992; 2003). Moreover, it was demonstrated that the PWI task can be modeled similarly to the classical color-word Stroop using the computational model WEAVER ++ (Roelofs, 2000; 2003). Specifically, the locus of the Stroop-like effect (i.e., the difference between congruent and incongruent trials) was linked to the lexical selection stage of the processing stream in both Stroop and PWI. Although intuitive, given the similarity between the color-word Stroop and the PWI tasks,

this approach remains a question of debate. Some researchers demonstrated pieces of evidence that suggest a rather earlier locus of PW interference, linked to perceptual encoding and conceptualization (Ayora et al., 2011; Dell'Acqua, Job, Peressotti, & Pascali, 2007; Dell'Acqua, Sessa, Peressotti, Mulatti, Navarrete, & Grainger, 2010; Van Maanen, Van Rijn, & Borst, 2009). In Chapters 2 and 3 of this dissertation I present research that discusses in detail this alternative account and gives experimental evidence for the word-planning locus of the PWI effect. Some other researchers, on the contrary, found no evidence for the PWI effect in perceptual encoding and conceptualization, as well as word planning, and conclude that the PWI effect is associated with articulatory buffer. Evaluation of this alternative account is outside the scope of this dissertation (for recent experimental data and evaluation of this approach, see Finkbeiner & Caramazza, 2006; Janssen, Schirm, Mahon, & Caramazza, 2008; Piai, Roelofs, Schriefers, 2011, 2012; Piai, Roelofs, Jensen, Schoffelen, Bonnefond, 2014; Roelofs, Piai, Schriefers, 2013).

1.2. ADJUSTMENT IN LANGUAGE PRODUCTION

Although the functional locus of PW interference remains a question of debate, until recently no account suggested that it might not be fixed to a certain stage of stimulus processing, but vary depending on certain manipulations in a task. However, Van Maanen and Van Rijn (2010) presented experimental evidence that the locus of PW interference can shift from early (i.e., perceptual encoding and conceptualization) to late (i.e., word planning) within a PWI task, depending on the previous trial type. They hypothesized that on post-incongruent trials participants exert more top-down control and postpone processing of a distractor word. This temporary adjustment shifts the functional locus of the PWI effect to the word planning stage. Alternatively, following a congruent trial, participants exert relatively little control and process the distractor word simultaneously with the picture, and the locus of the PWI effect remains early, linked to perceptual encoding and conceptualization (as they suggested in a previous work; Van Maanen et al., 2009). I discuss this issue and present our experimental evidence against such shift of functional locus of interference in Chapter 3 of this dissertation.

Control adjustments in Stroop-like tasks are usually demonstrated using the so-called Gratton effect (Gratton, Coles, & Donchin, 1992): The size of the Stroop-like effect (e.g., in RTs or error rates) is smaller following incongruent trials than following congruent trials. Outside the language domain the Gratton effect is usually interpreted in terms of the conflict-monitoring hypothesis (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Similarly, to the previously outlined interpretation by Van Maanen and Van Rijn (2010), the conflict-monitoring account states that the observed difference in the Stroop-like effect between post-congruent and post-incongruent trials is due to conflict adaptation, i.e. enhanced control exerted on

post-incongruent trials because of previously experienced conflict. Therefore, the conflict-monitoring account posits that the control adjustment effects (e.g., the Gratton effect) are driven by incongruent trials. This approach was further used to explain the Gratton effect in language production by a few groups of researchers (Duthoo, Abrahamse, Braem, Boehler, & Notebaert, 2014; Freund, Gordon, & Nozari, 2016; Van Maanen & Van Rijn, 2010). However, in a series of experiments employing linguistic and non-linguistic tasks, Lamers and Roelofs (2011) demonstrated that conflict adaptation cannot account alone for the multiple ways the Gratton effect manifests itself in RTs and error rates. They suggested an alternative account. First, control adjustments occur due to the fact that the amount of processing of the distractor changes depending on the previous trial type. More specifically, the attention window widens following congruent trials and narrows down following incongruent trials. Second, all post-incongruent trials (independently on the current-trial type) demonstrate increased RT compared to post-congruent trials, due to caution (to a various degree). We evaluate the conflict-monitoring and the two-factor accounts to language production in a series of experiments in Chapter 4.

Adjustments in top-down control depending on whether a congruent or an incongruent PWI trial is expected is only one of many situations in which a speaker anticipates an upcoming task and strategically modifies his or her behavior in order to perform well. More generally speaking, a speaker needs to allocate a certain amount of resources for speaking depending on the difficulty of the task. If the situation requires the speaker to produce an elaborate description of a certain object ("a big red cup"), it will likely take more processing capacity than if a simple reference is enough to communicate the idea ("a cup"). In Chapter 5 I describe electrophysiological patterns of allocation and use of processing capacity in phrase production.

1.3. OUTLINE OF THE DISSERTATION

This dissertation includes four studies that addressed various questions regarding competition and adjustment effects in language production. Chapter 2 presents an electroencephalography (EEG) study on the functional locus of the Stroop-like effect in color and picture naming. It was motivated by the discussion in the literature whether the functional locus of the Stroop-like effect in PWI is the same as in the classical color-word Stroop task (i.e., in word planning), or whether it is rather linked to earlier stages of the processing stream (i.e., perceptual encoding and conceptualization). Chapter 3 continues on the research presented in Chapter 2 and investigates the PWI effect from the point of view of control adjustments using EEG. The remaining two chapters follow up on adjustment effects in language production. Chapter 4 specifically addresses the question whether control adjustments are driven by conflict, as held by the conflict-monitoring account, or whether the pattern of control adjustments can

be better explained by expectancy of congruent trials. Furthermore, Chapter 4 investigates control adjustment effects in non-discrete responses (i.e., phrases), for the first time in the literature. The topic of adjustment effects in phrase production was researched further in Chapter 5, that used EEG in order to assess effects of complexity and switching.

CHAPTER 2

USING BRAIN POTENTIALS TO FUNCTIONALLY LOCALIZE STROOP-LIKE EFFECTS IN COLOR AND PICTURE NAMING: PERCEPTUAL ENCODING VERSUS WORD PLANNING

A slightly modified version of this chapter was published as:

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M, 2016 Using brain potentials to functionally localise Stroop-like effects in colour and picture naming: Perceptual encoding versus word planning. *PloS One*, *11*, 1-16.

ABSTRACT

The color-word Stroop task and the picture-word interference task (PWI) have been used extensively to study the functional processes underlying spoken word production. One of the consistent behavioral effects in both tasks is the Stroop-like effect: The reaction time (RT) is longer on incongruent trials than on congruent trials. The effect in the Stroop task is usually linked to word planning, whereas the effect in the PWI task is associated with either word planning or perceptual encoding. To adjudicate between the word planning and perceptual encoding accounts of the effect in PWI, we conducted an EEG experiment consisting of three tasks: a standard color-word Stroop task (three colors), a standard PWI task (39 pictures), and a Stroop-like version of the PWI task (three pictures). Participants overtly named the colors and pictures while their EEG was recorded. A Stroop-like effect in RTs was observed in all three tasks. ERPs at centro-parietal sensors started to deflect negatively for incongruent relative to congruent stimuli around 350 ms after stimulus onset for the Stroop, Stroop-like PWI, and the Standard PWI tasks: an N400 effect. No early differences were found in the PWI tasks. The onset of the Stroop-like effect at about 350 ms in all three tasks links the effect to word planning rather than perceptual encoding, which has been estimated in the literature to be finished around 200-250 ms after stimulus onset. We conclude that the Stroop-like effect arises during word planning in both Stroop and PWI.

2.1. INTRODUCTION

A key component of spoken language production is the retrieval of lexical information from long-term memory so that appropriate words can be planned and articulated. The process of word planning has been intensively investigated over the past few decades, resulting in detailed computational models of both normal and impaired word production (e.g., Dell, Schwartz, Nozari, Faseyitan, & Coslett, 2013; Levelt, Roelofs, & Meyer, 1999; Rapp, Goldrick, 2000; Roelofs, 1992; 2014; Ueno, Saito, Rogers, & Lambon Ralph, 2011). One of the workhorses in the study of spoken word production is the picture-word interference (PWI) task (e.g., Damian & Martin, 1999; Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Rayner & Springer, 1986; Schriefers, Meyer, & Levelt, 1990), which is often seen as an adaptation of the classic color-word Stroop task (e.g., MacLeod, 1991). In these tasks, speakers name pictures or colors while trying to ignore distractor words. Findings obtained with the PWI task have been taken to provide information about the underlying word planning process, not only in healthy adult speakers (e.g., Levelt et al., 1999) but also in aphasic patients with production impairments as a consequence of stroke (e.g., Hashimoto & Thompson, 2010) or neurodegenerative disease (e.g., Thompson et al., 2012). Moreover, the PWI task is employed in studies of word production in developmental language impairment (e.g., Seiger-Gardner & Schwartz, 2008).

A main assumption underlying the use of the PWI task in these studies is that it taps into word planning processes (e.g., Levelt et al., 1999) rather than pre-planning processes such as perceptual and conceptual encoding of the picture or post-planning processes such as articulatory buffering and the initiation of articulation. A word-planning locus has also been assumed for the Stroop task (e.g., Roelofs, 2003). However, this word-planning assumption has recently been contested by researchers who argue that PWI effects may have a pre-planning locus (e.g., Dell'Acqua, Job, Peressotti, & Pascali, 2007; Van Maanen, Van Rijn, & Borst, 2009) or a post-planning locus (e.g., Finkbeiner & Caramazza, 2006; Janssen, Schirm, Mahon, & Caramazza, 2008). Dell'Acqua et al. (2007) and Van Maanen et al. (2009) assumed a pre-planning locus for the PWI task but a word-planning locus for the Stroop task. The aim of the present article is to address the issue of the functional locus of distractor word effects in picture naming (PWI) and color naming (Stroop) by exploiting the ability of event-related brain potentials (ERPs) to provide information about the time course of effects between stimulus onset and articulation onset.

The chapter is organized as follows. We first describe the PWI and Stroop effects and the word planning account in somewhat more detail. Moreover, we discuss the findings that have been advanced against a word-planning locus in PWI and which have motivated an alternative pre-planning account (Dell'Acqua et al., 2007; Van Maanen et al., 2009). Next,

we outline a new ERP study that was designed to test between the pre-planning and word-planning accounts (for a recent evaluation of the post-planning account, see Roelofs, Piai, & Schriefers, 2013). Then, the results of the ERP study are reported. Finally, we discuss the consequences of our findings for the debate about the locus of distractor word effects in picture and color naming.

Since the PWI task was first introduced (Rosinski, Golinkoff, & Kukish, 1975), it has been taken to be a version of the color-word Stroop task (MacLeod, 1991). Both tasks employ compound stimuli with two dimensions, a nonlinguistic target dimension (a picture or a color) and a word as distractor dimension. Moreover, both tasks require a response to the target dimension while trying to ignore the distractor dimension (i.e., naming the picture or color while ignoring the word). Studies on color or picture naming usually employ stimuli with dimensions that are incongruent (e.g., the word *blue* written in red ink; the word *dog* superimposed on a picture of a cat) and congruent (e.g., the word *red* written in red ink; the word *cat* superimposed on a picture of a cat) or neutral/unrelated (e.g., a neutral series of Xs written in red ink or superimposed on a picture of a cat; the unrelated word *house* superimposed on a picture of a cat, e.g., MacLeod, 1991).

In general terms, the processing stream for a single color or picture naming trial consists of three main stages. First the color or the picture is perceived and conceptually identified based on the stimulus features, henceforth the *perceptual encoding* stage (cf. Van Maanen et al., 2009). Then the corresponding spoken word is planned based on information retrieved from long-term lexical memory, henceforth the *word planning* stage (cf. Levelt et al., 1999; Roelofs, 2003; 2014). Finally, the planned spoken name is articulated (e.g., Levelt et al., 1999; Roelofs, 2014). In Stroop or PWI, depending on the condition, the distractor word either interferes with this target processing stream or facilitates it (MacLeod, 1991). One of the most consistent behavioral findings in the Stroop and PWI tasks is the Stroop-like effect: The reaction time (RT) on incongruent trials is longer than on congruent trials (e.g., Glaser & Düngelhoff, 1984; Glaser & Glaser, 1982). The Stroop-like effect in PWI is often taken to include a semantic interference effect (i.e., the RTs on incongruent trials are longer than on unrelated trials), and an identity facilitation effect (i.e., the RTs on congruent trials are shorter than on unrelated trials; Glaser & Düngelhoff, 1984).

According to the word-planning account, the distractor word may compete with the picture/color name at the stages of lemma retrieval and word-form encoding, jointly referred to as word planning by Levelt et al. (1999) and Roelofs (2003; 2014). The specific interference effects associated with these stages were discussed by Roelofs (2014). In particular, a congruent distractor helps lemma retrieval and word-form encoding for the target name because activation from the picture/color and distractor word converges on the target lemma and

word form. In contrast, an incongruent distractor hampers lemma retrieval and word form encoding, because picture/color and distractor word activate competing lemmas and word forms (for a computational implementation of this view, see Roelofs, 2003). Importantly, on this view, the way a distractor word helps or hampers the planning of the target word (i.e., picture or color name) does not differ between the Stroop and PWI tasks. In contrast, according to the perceptual-encoding account of Van Maanen et al. (2009), the speed of perceptual encoding differs between pictures and colors, which should yield a difference in the locus of interference in Stroop vs. PWI.

In combination with estimates of the time windows of the main stages of word production from meta-analyses (Indefrey & Levelt, 2004; Indefrey, 2011), ERPs can show at what stage of processing distractors interact with target processing during Stroop or PWI trials. A modulation of the N400 is consistently reported as a neurophysiological signature of the Stroop-like effect for both tasks (e.g., Liotti, Woldorff, Perez, & Mayberg, 2000, Badzakova-Trajkov, Barnett, Waldie, & Kirk, 2009, for Stroop; Piai, Roelofs, Van der Meij, 2012; Xiao, Zhang, Jia, Zhang, & Luo, 2010, for PWI). In particular, ERPs over centro-parietal electrodes typically deflect more negatively on incongruent trials than on congruent trials between approximately 300 ms and 500 ms after stimulus onset. The timing of this effect corresponds to the word planning rather than the perceptual encoding stage. Perceptual encoding has been estimated to be completed around 200-250 ms after stimulus onset (Indefrey & Levelt, 2004; Indefrey, 2011), thus before the onset of the N400 effect (which is around 300-350 ms). However, one PWI study obtained a semantic effect not only in the N400 time window but also earlier, at about 100 ms (Dell'Acqua et al., 2010), although other studies found no such early effect (e.g., Piai et al., 2012; Blackford, Holcomb, Grainger, & Kuperberg, 2012). The observation that the Stroop-like effect in both Stroop color naming and PWI is reflected in the N400 suggests that word planning rather than perceptual encoding is the locus of the effect in both tasks (cf. Roelofs, 2003), whereas the early semantic interference effect in PWI by Dell'Acqua et al. may suggest different loci in the Stroop task and the PWI task, as maintained by Van Maanen et al. (2009).

The assumption of a word-planning locus of the Stroop-like effects in both Stroop and PWI has also been contested based on RT effects obtained by Dell'Acqua et al. (2007) using a dual-task design (cf. Pashler, 1984), concentrating on the semantic interference effect (see Figure 2.1). On each trial in the study of Dell'Acqua et al. a tone stimulus was followed by a PWI stimulus with a certain stimulus onset asynchrony (SOA) and participants had to respond to both stimuli in correct order (i.e., first they had to indicate the pitch of the tone by a manual response and then they had to name the picture). Naming RTs in the PWI task were longer at short than at long SOAs. More importantly, Dell'Acqua et al. observed that the magnitude of the semantic interference effect in picture naming was smaller at short than at long SOAs.

Thus, the semantic interference effect and the SOA effect were underadditive. Assuming a response-selection bottleneck (Pashler, 1994), the underadditivity of effects suggests that the semantic interference effect arises before word planning in picture naming (i.e., in perceptual encoding; (Figure 2.1B). In contrast, using a similar dual-task design, Fagot and Pashler (1992) observed that the magnitude of the Stroop-like effect in color naming did not differ between long and short SOAs, suggesting that the effect arises in word planning in color naming (Figure 2.1A). Given this difference in effects between PWI (underadditivity) and Stroop (additivity), Dell'Acqua et al. concluded that there is a different locus of the Stroop-like effects in Stroop and PWI (i.e., word planning vs. perceptual encoding, respectively). Based on computer simulations using ACT-R and RACE/A, Van Maanen et al. (2009) argued that the locus of the Stroop and PWI effects is different (i.e., word planning vs. perceptual encoding) because of a difference in the speed of perceptual encoding between colors and pictures.

However, subsequent experimental research on the semantic interference effect in the PWI task using the dual-task design showed diverging results. On the one hand, the semantic interference effect and the SOA effect were reported to be underadditive by Ayora et al. (2011) and Van Maanen, van Rijn, and Taatgen (2012), supporting a perceptual encoding locus of the semantic interference effect in PWI. On the other hand, the semantic interference effect and the SOA effect were reported to be additive by Schnur and Martin (2012) and Piai, Roelofs, and Schriefers (2014), which speaks in favor of a word planning locus of the semantic interference in PWI. Moreover, Piai et al. observed that the magnitude of the Stroop-like effect did not differ between SOAs for both PWI and Stroop (i.e., the Stroop-like and SOA effects were additive in both tasks) using a single group of participants performing both tasks. This similarity in RT effects between tasks is in line with the observation from ERP studies that the Stroop-like effect in both Stroop and PWI is reflected in the N400, suggesting that word planning rather than perceptual encoding is the locus of the effect in both tasks (e.g., Aristei, Melinger, & Abdel Rahman, 2011; Liotti et al., 2000; Piai et al., 2012). Moreover, Piai, Roelofs, and Schriefers argued that different groups of participants may adopt different processing strategies in dual-task performance (cf. Meyer & Kieras, 1997), which may explain the diverging results (i.e., additive vs. underadditive effects).

To summarize, extant studies report inconsistent results and entail different loci of semantic interference for the PWI task: either early, linked to the perceptual encoding stage (Dell'Acqua et al., 2007; Dell'Acqua et al., 2010; Van Maanen et al., 2009), or later, linked to word planning (Schnur & Martin, 2012; Piai, Roelofs, & Schriefers, 2014). In contrast, the evidence concerning the Stroop-like effect in color naming is consistent (Fagot & Pashler, 1992; Piai, Roelofs, & Schriefers, 2014), and suggests a locus of the effect in word planning.

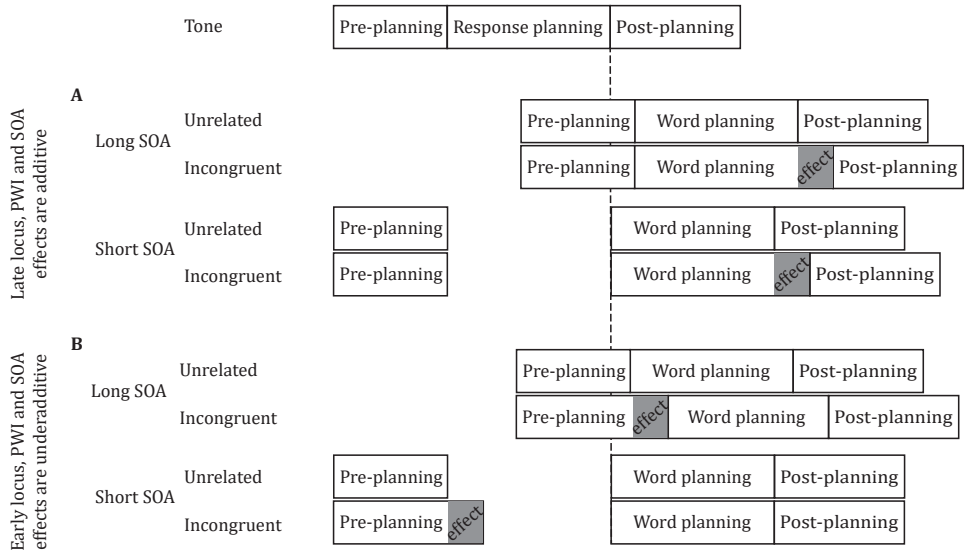

FIGURE 2.1. An illustration of the response bottleneck account. In a dual task, planning of response to the second stimulus is postponed until the response to the first stimulus is selected. For example, in a dual task with tone discrimination as Task 1 and PWI as Task 2, word planning cannot be started before a response to a tone is selected. The PWI effect (or the Stroop-like effect) is taken as a difference between RTs on incongruent trials and RTs on unrelated (or congruent) trials. **A.** If the effect is linked to the word-planning stage of the processing stream, no difference in the effect size is expected between long SOA and short SOA conditions (i.e., PWI and SOA effects are additive). **B.** However, if the effect is linked to perceptual and conceptual encoding (i.e., the pre-planning stages of the processing stream), the effect is expected to be smaller on short-SOA trials than on long-SOA trials. This underadditivity of PWI and SOA effects is due to the fact that the PWI effect is absorbed into slack created by the response bottleneck.

Because of the discrepancy in empirical results, researchers still have not found agreement on the functional locus of Stroop-like effects in the PWI and color-word Stroop tasks (e.g., Dell'Acqua et al., 2007; Roelofs, 2003; 2014; Van Maanen et al., 2009). One of the main reasons why it has not been possible to adjudicate between the perceptual encoding and the word planning accounts is that the data sets used for comparisons between Stroop and PWI are usually from different studies, which means that different participant groups and sometimes different experimental parameters, stimulus type sets, and even experimental contrasts are involved. For instance, Dell'Acqua et al. compared the underadditivity of the semantic interference effect and the SOA effect that they obtained in their PWI study with the additivity of the Stroop-like effect and the SOA effect reported by Fagot and Pashler (1992) to motivate their conclusion of a different locus of the effects in the Stroop and the PWI tasks. Likewise, the computational study of Van Maanen et al. involved a direct comparison between

the dual-task data of Fagot and Pashler (a Stroop-like effect in the Stroop task performed by American participants) and the dual-task data of Dell'Acqua et al. (a semantic interference effect in the PWI task performed by Italian participants). The difference in participant groups (American vs. Italian) and experimental contrasts (semantic vs. Stroop-like) seems problematic, given that Piai, Roelofs, and Schriefers (2014) observed that the magnitude of the Stroop-like effect did not differ between SOAs for both PWI and Stroop using a single group of (Dutch) participants performing both Stroop and PWI tasks. As indicated, Piai et al. argued that different participant groups may adopt different processing strategies in dual-task performance, explaining the discrepancy in empirical results between studies.

The problem with direct comparisons of the results of different studies not only holds for the RT findings from dual-task studies, but also for the EEG studies on PWI and Stroop (e.g., Badzakova-Trajkov et al., 2009; Dell'Acqua et al., 2010; Liotti et al., 2000; Piai et al., 2012; Xiao et al., 2010). To make a convincing case in favor or against a common locus of the Stroop-like effect in PWI and Stroop, the ERP effects should be examined for a single group of participants performing both tasks. Moreover, to properly compare the effects elicited by the two tasks, experimental parameters have to be highly comparable between tasks. In addition, to relate the ERP modulations to processing stages, it is desirable that RTs are comparable between tasks (cf. Indefrey & Levelt, 2004; Indefrey, 2011). This is not the case for the typical Stroop versus PWI comparison. RTs in the PWI task are usually longer than RTs in the Stroop task, a difference that can be attributed to the fact that Stroop experiments use fewer different to-be-named stimuli (usually three colors) than PWI studies (usually 20 pictures or more). In the following, we report an ERP study testing a single group of participants on standard color-word Stroop (three colors and three distractor words), Standard PWI (39 pictures and 39 distractor words), and Stroop-like PWI (three pictures and three distractor words).

2.2. METHOD

2.2.1. Ethics Statement

The study was performed within the line of research approved by the Ethics Committee of the Faculty of Social Sciences at Radboud University, Nijmegen (ECG-2013-2504-095-102) and followed the World Medical Association Declaration of Helsinki. All participants signed an informed consent form before the start of the experiment.

2.2.2. Participants

Twenty-eight participants (19 to 28 years old, mean 24 years, six male) took part in the study for monetary compensation. All participants were right-handed, native speakers of Dutch, with normal or corrected to normal vision and no history of neurological disorders. The data of four participants had to be excluded from analysis due to a large number of artifacts in the EEG recording.

2.2.3. Procedure

The experiment consisted of three tasks: the Stroop task, the Standard PWI task, and a Stroop-like version of the PWI task. The order of the tasks was randomized across participants, and a short break was offered to the participants between the tasks. The procedure was identical for all three tasks. The participant was seated approximately 70 cm away from a computer screen. She first read the instruction on the screen and familiarized herself with the stimulus set, then performed a practice session of eight trials, after which she performed three blocks of the task (78 trials per block) separated with short breaks.

The stimuli were presented in the center of the screen, on a black background. For the Stroop task, the participants were presented with a color word printed in blue (RGB decimal code 0,102,204), red (255,51,0), or green (51,153,0). For the two versions of PWI, a line drawing (4-by-4 cm) with a superimposed distractor word was presented. In all three tasks the words were printed in low-case Arial 24pt letters.

The stimuli were randomized using Mix software (Van Casteren & Davis, 2006) in such a way that each participant had an individual order of stimulus presentation. Each task consisted of 234 trials, half of which were congruent and the other half were incongruent. For Standard PWI, 39 pictures paired with congruent and incongruent distractor words were presented once in each block. For the Stroop task, three colors were presented 13 times in either condition (congruent or incongruent color word) in each block. For Stroop-like PWI, three pictures were presented 13 times in either condition in each block. A miniblock-based pseudo-randomization was applied to the Stroop-like PWI and Standard PWI stimuli. The size of a mini-block was 6 trials (three pictures/colors in two conditions). The stimuli were pseudo-randomized within mini-blocks and the mini-blocks were further concatenated to form stimulus lists for blocks. In this way we ensured that stimuli that employed the same picture/color were evenly distributed throughout the block and the task in general. A trial was identical for all three tasks. It started with a presentation of four asterisks for 1200 ms to mark the blinking time for the participant. This was followed by a blank interval varying

randomly between 1000 and 1300 ms, during which the participant was instructed to fixate on the center of the screen (though no actual fixation mark was present). For the next 750 ms a stimulus was presented. After offset of the stimulus, the screen remained blank for 700 ms. The participant was instructed to respond as quickly and accurately as possible. The total duration of an experimental session was about 50 minutes.

2.2.4. Materials

The response set for the Standard PWI task consisted of 13 groups of three co-hyponyms (see Table A1) drawn from semantic categories that are usually used in PWI experiments (e.g., birds, tools, fruits, means of transport). The Stroop-like PWI response set consisted of responses from a single semantic category (i.e., animals), which was not part of the Standard PWI response set. The Stroop task response set consisted of three color names (red, green, and blue). In order to obtain an incongruent stimulus, a color or a picture was paired with one of the other two color or picture names from the same co-hyponym group (see Table A1), and for a congruent stimulus a color or a picture was paired with the name of this color or picture.

The number of responses in the Stroop and the Stroop-like PWI tasks was limited to three for the following reasons. All studies from the literature cited in the current chapter employed three or four colors. The studies most directly relevant for the current study (i.e., Fagot & Pashler, 1992; Piai, Roelofs, & Schriefers, 2014) employed three colors. Moreover, an experimental aim of the current study was to amplify the difference between the Standard PWI task, on the one hand, and the Stroop and the Stroop-like PWI tasks, on the other hand, with respect to stimulus set size. Thus, three was the lowest possible number of colors that allowed for a good randomization while preventing strict alternations or many repetitions.

The distractor words for the PWI tasks were common-level nouns. The word frequency was estimated with the online version of SUBTLEX-NL (http://crr.ugent.be/isubtlex/) for all stimuli. The variance within the groups was kept as low as possible. The word frequency estimates per million words are available in Table A1.

Line drawings were partially taken from the picture database of the Max Planck Institute for Psycholinguistics (some of them were further edited). The rest was taken from open source internet resources or drawn from scratch. The materials were preliminarily tested in a pilot experiment to assure untroubled naming of the color or picture.

For the pilot experiment, we used three colors for the Stroop task, three pictures for the Stroop-like PWI task, and 45 pictures for the Standard PWI task. Twelve participants were instructed to name the color or the picture. High name agreement among subjects was

observed. Information about problematic items (i.e., among the 45 pictures) in the stimuli list was obtained by analyzing the types of errors the participant made (e.g., using a different than expected name for a picture in the incongruent condition) or via informal feedback of the participant.

2.2.5. Recording

The overt responses were recorded for offline reaction time measurements with PRAAT (Boersma, 2001). Errors in naming were detected during the experiment and the corresponding trials were excluded from the analysis.

The EEG setup consisted of 64 active electrodes embedded in a 10-20 international system electrode cap (ActiCAP 64Ch Standard-2, Brain Products), and the data were online referenced to FCz and re-referenced offline to the average of both mastoids. Additionally, three pairs of passive bipolar electrodes were used in order to register eye and lip movements. The vertical EOG was recorded from the electrodes placed above and below the left eye. Two electrodes were placed on the temples in order to record the horizontal EOG. The lip movements (EMG) were registered with the electrodes placed at the left orbicularis oris superior and the right orbicularis oris inferior. The data were recorded at a sampling rate of 1000 Hz.

2.2.6. Analysis

The data analysis was performed with Fieldtrip (Oostenveld, Fries, Maris, Schoffelen, 2011) and custom analysis scripts using Matlab v.8.1.0.604 (R2013a, The MathWorks, Inc.). Trials were defined from 500 ms before the onset of the stimulus until 100 ms before the overt articulation onset, so the length of epochs differed from trial to trial. Then the EOG channels were visually inspected for artifacts and those trials that contained blinks or eye movements were completely removed from the analysis. The EEG data were further inspected for muscle artifacts, which could be seen as high-frequency noise with relatively sudden onset at about 100 - 200 ms before the speech onset, occurring over a spatially distributed range of EEG channels. Trials with muscle artifacts occurring before 300 ms post-stimulus onset were discarded completely. Otherwise the trials were truncated until just before the onset of the muscle artifact. As a result, the epochs were of different lengths for different trials (and, naturally, for different participants and tasks).

One participant's data were excluded from the analysis due to the extremely early onset of muscular artefacts in the Stroop-like PWI task (approximately 75% of her trials were shorter than 400 ms post-stimulus onset following the procedure of muscular artifacts rejection).

In total, due to errors in naming, eye movements or blinks, and early onset of EMG activity 11.2% of trials were rejected in the Stroop task, 9.4% of trials were rejected in the Stroop-like PWI task, and 13.1% of trials were rejected in the Standard PWI task. The median number of the trials rejected per participant was 22 (Q1=16.25, Q3=32.75, IQR=16.5) for the Stroop task, 20 (Q1=11.25, Q3=30.25, IQR=19) for the Stroop-like PWI task, and 24 (Q1=20.25, Q3=45.25, IQR=25) for the Standard PWI task.

The artifacts-free data were further band-pass filtered at 0.2-40 Hz using a onepass-zerophase linear non-causal hamming-windowed FIR filter (transition width 0.4 Hz, stopband attenuation -53 dB, maximal passband deviation 0.22%). The data were further baselined using a pre-stimulus interval (from 300 ms pre-stimulus onset to stimulus onset).

The ERP analysis was performed by averaging waveforms across trials per condition and per participant. Because of the variable lengths of the individual trials, the ERPs for the different participants, conditions and tasks were of variable length. Also, for any participant, condition and task, the number of trials contributing to the average ERP diminished as a function of time after stimulus onset, due to differences in response latency. Statistical tests were applied to time windows in which all participants contributed data. Thus, the length of the shortest ERP across participants and conditions defined the time-point that limited the between-condition comparison (i.e., incongruent vs. congruent) within each task. This time-point was 530 ms post-stimulus onset for the Stroop task, 563 ms post-stimulus onset for the Stroop-like PWI task, and 781 ms post-stimulus onset for the Standard PWI task. The median number of trials included in the ERPs at every time-point is presented in Figure A1.

The resulting ERPs were submitted to within-participants cluster-based permutation tests (Maris & Oostenveld, 2007) in order to assess statistical significance of the difference between conditions for each task, which was operationalized as follows. First, for every electrode/time-point a paired-samples *t*-value was calculated. The main comparison between congruent and incongruent conditions was based on all time-points from 300 ms pre-stimulus onset and up to the end of the shortest individual ERP waveform over participants and conditions. Another comparison was intended to test specifically for early effects in the Stroop-like PWI task and the Standard PWI task. For this latter comparison only the time-points from the stimulus onset until 300 ms post-stimulus onset entered the analysis. The paired-samples *t*-values that were larger than ±2.7 (which corresponded to the 97.5th quantile for a two-sided test) were selected and the electrode/time-points associated with these *t*-values were clustered based on spatial and temporal adjacency. The spatio-temporal clusters were defined to contain at least two spatial neighbors (based on the EEG electrodes layout) and one time sample (1 ms given the sampling frequency of 1000 Hz). A sum of the *t*-values within the cluster served as a cluster statistic. A permutation distribution under the null-hypothesis of exchangeability

across conditions was constructed by random re-assignment of the condition labels to the original individual ERPs, followed by the construction of spatio-temporal clusters, in the same way as for the observed data. We used 1000 permutations and determined a cluster-based *p*-value, as the proportion of random permutations that yielded a larger cluster statistic than the cluster in the original data. In case the *p*-value appeared to be smaller than the critical alpha-level (0.025 for the two-sided test), it was concluded that the experimental conditions were significantly different.

2.3. RESULTS

2.3.1. Behavioral results

In all three tasks, participants gave faster and more accurate responses in the congruent condition than in the incongruent condition. Table 2.1 gives the mean RTs and error percentages per task and condition.

TABLE 2.1. Performance per condition for Stroop, Stroop-like PWI and Standard PWI

	Stroop			**Stroop-like PWI**			**Standard PWI**		
condition	RT	SD	ER	RT	SD	ER	RT	SD	ER
congruent	565	84	0.8	589	69	0.7	677	80	1.4
incongruent	658	96	5.1	696	69	3.7	831	71	7.3

RT = response time (in milliseconds), SD = standard deviation (in milliseconds), ER = error rate (in percent), PWI = picture-word interference.

For the RT data, a 3 x 2 repeated-measures ANOVA with the factors task (Stroop, Stroop-like PWI, Standard PWI) and condition (congruent, incongruent) showed significant main effects of task ($F(2,44) = 50.9$, $p < .001$), condition ($F(1,22) = 281.1$, $p < .001$), and an interaction between task and condition ($F(2,44) = 31.3$, $p < .001$). Further analysis comparing RTs in the Stroop and the Stroop-like PWI tasks showed an effect of condition ($F(1,22) = 242.4$, $p < .001$) and a marginally significant effect of task ($F(1,22) = 4.31$, $p = .0497$), while the interaction between task and condition ($F(1,22) = 4.15$, $p = .054$) did not reach significance. Comparison of RTs in the Stroop-like PWI and the Standard PWI tasks showed significant main effects of task ($F(1,22) = 77.5$, $p < .001$), condition ($F(1,22) = 305.9$, $p < .001$), and an interaction between task and condition ($F(1,22) = 35.7$, $p < .001$). The latter reveals that the Stroop-like effect is larger for the Standard PWI than the Stroop-like PWI task.

For the accuracy data, a 3 x 2 repeated-measures ANOVA with the factors task and condition showed significant main effects of task ($F(2,44) = 9.2$, $p < .001$), condition ($F(1,22) = 39$, $p < .001$), and an interaction between task and condition ($F(2,44) = 4.09$, $p = .024$). Further analysis comparing error rates in the Stroop and the Stroop-like PWI tasks showed an effect of condition ($F(1,22) = 24.3$, $p < .001$), while the effect of task ($F(1,22) = 3.3$, $p = .084$) and the interaction between task and condition ($F(1,22) = 3.6$, $p = .071$) did not reach significance. Comparison of error rates in the Stroop-like PWI and the Standard PWI tasks showed significant main effects of task ($F(1,22) = 17.1$, $p < .001$), condition ($F(1,22) = 42.8$, $p < .001$), and an interaction between task and condition ($F(1,22) = 6.57$, $p = .018$).

2.3.2. ERP results

Group average ERPs for the three tasks and the two conditions are shown in Figure 2.2A. Inspecting the curves visually, it can be observed that the waveforms deflected more negatively in the incongruent condition than in the congruent condition in the N400 time window but no earlier than 350 ms after stimulus onset in all three tasks.

We aimed to test between the perceptual encoding and the word planning accounts of the locus of the Stroop-like effect. Whereas the word-planning account predicts N400 effects for all three tasks, the perceptual encoding account does not make clear claims about the nature and directionality of the predicted early effect in the PWI tasks. Given this, we performed a two-sided cluster-based permutation test on the entire ERPs and including all electrodes. This yielded significant differences between congruent and incongruent stimuli in all three tasks. In the Stroop task a negative centro-parietal cluster was observed from 345-460 ms post-stimulus onset ($p = .004$). For the Stroop-like PWI task and the Standard PWI tasks negative fronto-centro-parietal clusters were observed (Stroop-like PWI: from 362-458 ms post-stimulus onset, $p = .008$; Standard PWI: from 328-571 ms post-stimulus onset, $p = .004$). The topographies of the effects are shown in Figure 2.2B. The ERPs over all electrode sites are shown in Figure A2, the difference ERPs for all three tasks plotted using the same axes are shown in Figure A3.

Additional analyses were run on the early time-windows (from 0-300 ms post stimulus onset) in order to test for early ERP effects in the PWI tasks. Given that there were no clear expectations with respect to the direction of the possible effects, two-sided cluster-based permutation tests were used, but no significant differences were found for either of the tasks.

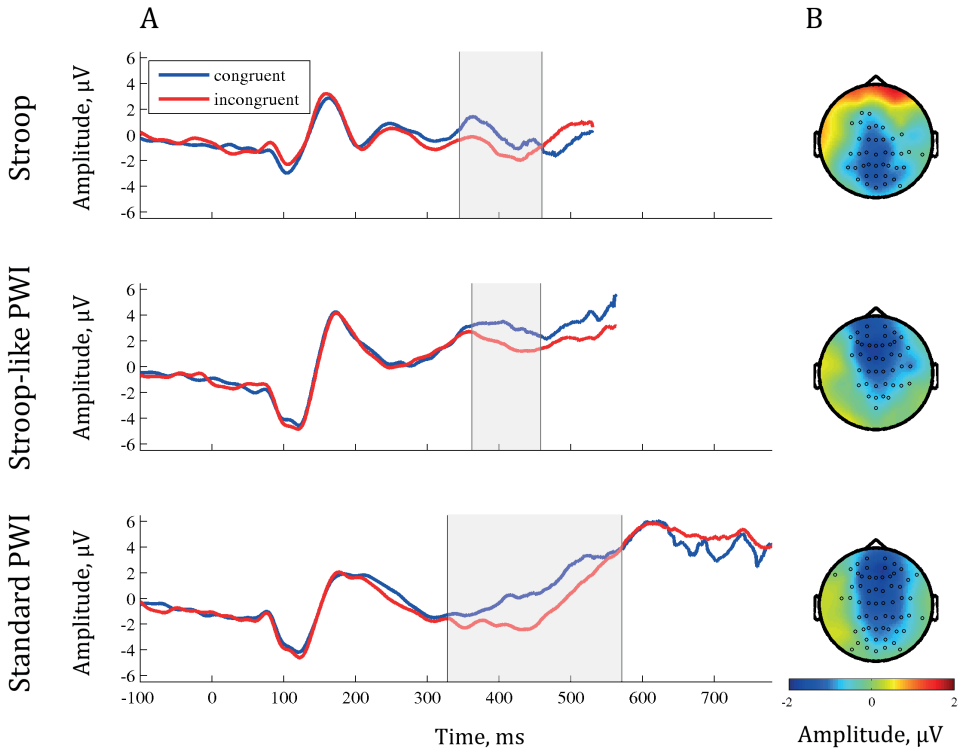

FIGURE 2.2. The N400 effects. **A.** Group-average sensor-level ERPs recorded from one site (Cz) in the Stroop task (top row), the Stroop-like PWI task (middle row) and the Standard PWI task (bottom row). The shaded area marks the time window in which the difference between conditions was significant. **B.** Topographies of the N400 effect (incongruent vs. congruent condition) in the Stroop, Stroop-like PWI, and Standard PWI tasks, respectively. The topographical distributions were calculated through averaging amplitudes of the difference group ERPs within the time-windows in which the difference between conditions was significant. The electrode sites that entered the spatio-temporal cluster based on which we rejected our null hypothesis are highlighted.

2.4. DISCUSSION

As discussed in the introduction, some previous studies have compared the Stroop-like effect in the Stroop and PWI tasks to assess the functional locus of the effect: perceptual encoding versus word planning. However, differences in design and participant groups have made direct comparisons between the Stroop and PWI tasks problematic. Therefore, conclusions based on these earlier results are also problematic. To remedy these problems, we conducted a study that involved a standard Stroop task (three colors) and a standard PWI task (39 pictures), as well as a Stroop-like PWI task (three pictures) within the same group of participants. The

within-participant design allowed us to directly compare color and picture naming in Stroop-like tasks with respect to the electrophysiological and behavioral effects. Moreover, inclusion of the Stroop-like PWI task, which yields RTs similar to the standard Stroop task, facilitated interpretation of the electrophysiological effects.

As concerns the RTs, we found the Stroop-like effect in all three tasks. The magnitude of the effect differed somewhat among tasks. The Stroop-like effect was approximately 100 ms in the Stroop and Stroop-like PWI tasks, and it was approximately 150 ms in the Standard PWI task. The overall mean RTs in the Stroop and Stroop-like PWI tasks were much shorter than the mean RT in the Standard PWI task (i.e., the RT difference was approximately 120 ms), whereas the difference in RTs observed between the Standard Stroop task and the Stroop-like PWI task was small (i.e., about 30 ms). This pattern of results shows that having the same number of stimuli (i.e., three) in the Stroop and PWI tasks helps to obtain similar RTs.

As concerns the ERPs, a significant difference in negativity between the incongruent and congruent conditions was observed in all three tasks within a time-window from about 350 to 500 ms post-stimulus onset, which is usually referred to as the N400 effect. However, visual inspection of the topographies of the condition effect (Figure 2.2B) and group-averaged ERP waveforms (Figure A2) revealed some differences in the N400 effect among the tasks. This raises the question whether really an N400 effect is obtained in all three tasks.

For the Stroop task, the difference between incongruent and congruent conditions had a rather posterior maximum, approximately at the Pz site, while for the two versions of PWI the effect had a fronto-central distribution. For the Stroop task, the post-central sites exhibited a clear negative-going wave within the time-window of the significant condition effect (but note the positive-going deflection over the Cz electrode in both conditions). For the Stroop-like PWI task, the ERPs show a negative-going potential for the incongruent condition and only a slight change for the congruent condition, which creates a resulting negative-going effect (i.e., an N400 modulation). For the Standard PWI task, as for the Stroop-like PWI task, the group-average ERPs exhibited negative-going deflections between 350 and 500 ms post-stimulus onset for the incongruent condition. One might argue that since for the majority of the central electrode sites this was the second negative peak over the waveform, the difference between conditions should be interpreted as an N2 effect. However, we interpreted this as an N400 effect. First, the peak was reached at approximately 450 ms for both PWI tasks, while the N2 component usually peaks at 250 – 300 ms after the stimulus onset (Folstein & Van Petten, 2008). Second, there exists only limited evidence that an N2 effect can be observed in the Stroop task (in its manual version) and no evidence that it is observed in an overt-production PWI task. According to Folstein and Van Petten, "in some studies the incongruity [N400] effect appears as an extension of the central N2 in time", "has a similar central scalp

distribution" and is "similarly sensitive to manipulations of cognitive control". However, in our study, in neither Stroop-like PWI nor Standard PWI did the N400 effect appear as an extension of an earlier effect, as evidenced by visual inspection and statistical analysis. To conclude, although the exact shapes of the waveforms differed slightly between tasks, closer inspection of all group-averaged ERP data supported the claim that the effect mostly corresponded to a negative-going deflection in the ERPs and occurred within the standard N400 time-window in all three tasks. Therefore, we concluded that the same N400 effect was observed in all three tasks.

The difference in the exact distribution of activity across time-points and EEG sites was expected, since the tasks utilized different stimuli, stimulus set sizes, and responses (i.e., words in the Stroop task, pictures in the two versions of the PWI task; small stimulus set in the Stroop and Stroop-like PWI tasks, large stimulus set in the Standard PWI task; color names in the Stroop task, object names in the PWI tasks). This may have yielded differences between tasks in stimulus processing at multiple stages from visual perception to response execution.

Kutas, Van Petten, and Kluender (2006) argued that different materials (e.g., pictures, printed words, spoken words, nonlinguistic sounds) yield different N400 wave shapes and topographies. Using the classic color-word Stroop task with overt spoken, covert spoken, and manual responding, Liotti and colleagues (2000) observed a common N400 difference between incongruent and congruent trials across task versions, but different corresponding wave shapes and topographies among versions. The same was observed by Donohue, Liotti, Perez III, and Woldorff (2012) using an auditory Stroop-like task with overt spoken, covert spoken, and manual responses. Thus, both stimulus-related and response-related differences in task design influence how the N400 effect is manifested. However, differences in wave shape and topographical distributions have not been taken as evidence that the source of the N400 effect is different among tasks. Instead, it is generally assumed that despite differences in wave shape and topography among tasks, there exists a general brain process generating the N400-like response (e.g., Holcomb & Anderson, 1993). Important for the question that we aimed to answer in our study, the difference between congruent and incongruent trials only appeared to be statistically significant within a particular time-window (i.e., the N400 window) for all three tasks (see Figure 2.2) over substantially overlapping centro-parietal spatio-temporal clusters of electrodes (see Figure A4).

While the difference between the incongruent and congruent conditions was consistently observed in the N400 time window, no earlier ERP effects were found for the PWI tasks, even with a pre-specified time-window of analysis that increased statistical sensitivity. With respect to the absence of an early ERP effect, it is important to compare our study to the study by Dell'Acqua et al. (2010), who reported a significant ERP effect in the time-window

from 50-250 ms post-stimulus onset. First of all, the design and experimental contrasts of the two studies were different. In the current study we investigated the Stroop-like effect and therefore compared incongruent (or, semantically related) and congruent (or, identity) trials, whereas Dell'Acqua et al. concentrated on the semantic interference effect and compared incongruent (or, semantically related) and neutral (or, semantically unrelated) trials. Second, the two studies employed substantially different methods of EEG data analysis. In our study we band-pass filtered the data at 0.2-40 Hz using a zero-phase linear non-causal filter. Dell'Acqua et al. used a lower cut-off for the high-pass filter (0.01 Hz) and a higher cut-off for the low-pass filter (80 Hz, unfortunately, filter characteristics are unknown), which might have been suboptimal for their data. Based on visual inspection of the ERP waveforms provided in their article, one might conclude that there were low-frequency drifts present in the data (both pre- and post-stimulus, see Figures 2 and 3 in Dell'Acqua et al.), which might have triggered the between-conditions differences. Furthermore, the two studies differed with respect to the statistical procedures used in ERP analysis in order to assess the differences between conditions. We employed a cluster-based permutation approach that assessed the differences between conditions using per-time-point per-electrode ERP data. Dell'Acqua et al. estimated condition-specific activity over nine predefined regions of interest (comprised of unequal numbers of electrodes) within two time windows, which were further compared with series of Bonferroni-corrected *t*-tests. Although both methods control for the false alarm rate, it is hard to compare these tests with respect to sensitivity. Finally, in the study by Dell'Acqua et al. only 12 participants (for the semantic-interference condition) were tested, while we analyzed the data of 23 participants. It should be noted that the absence of an early effect (i.e., before 300-350 ms after stimulus onset) in our study agrees with the findings of other studies in the literature (e.g., Aristei et al., 2011; Blackford et al., 2012; Piai et al., 2012; Piai, Roelofs, Jensen, Schoffelen, & Bonnefond, 2014).

The pattern of results observed in the current study supports the hypothesis that the Stroop-like effect is arising at the word-planning stage in both color and picture naming (Roelofs, 2003). Moreover, the results contradict the hypothesis of Van Maanen et al. (2009) that the Stroop-like effect has different loci in color and picture naming (i.e., word planning vs. perceptual encoding). The latter hypothesis would predict an early Stroop-like effect (i.e., before 200-250 ms after stimulus onset) in the Standard and Stroop-like PWI tasks, contrary to our findings (i.e., the effects started to occur at around 350 ms in all tasks).

Due to the analysis procedure used in the current study, namely, complete rejection of the epochs of EEG data that were contaminated by muscle artifacts, studying brain activity close to (or after) onset of articulation was not possible (but see Riès, Janssen, Burle, & Alario, 2013). For this reason, we cannot provide any evidence for or against the post-planning

account, which locates the Stroop-like effects in the Stroop and PWI tasks in an articulatory buffering stage (Finkbeiner & Caramazza, 2006; Janssen et al., 2008). The only ERP effect we obtained (the N400 effect) had an onset and offset within the word-planning time window.

Our conclusion that the Stroop-like effect in Stroop and PWI tasks arises in word planning rather than perceptual encoding does not necessarily generalize to other task situations and response modalities. In particular, based on findings obtained with the Stroop task and an arbitrary mapping of colors onto manual responses, De Houwer (2003) and Van Veen and Carter (2005) argued for both perceptual encoding and response planning loci of Stroop-like effects. De Houwer used a task with four colors and two response buttons such that two colors were mapped onto one response button and the other two colors were mapped onto the other response button. Color and word could be the same and thus require the same response button (congruent), they could be different but require the same response button (perceptual difference), or they could be different and also require different response buttons (perceptual and response difference). RTs were longer when there was both a perceptual and a response difference than when there was a perceptual difference only, suggesting a locus of the Stroop-like effect in word planning. Moreover, RTs were shorter when color and word were the same than when they were different but required the same response, suggesting a locus in perceptual encoding. It is unclear, however, whether these results obtained with arbitrary stimulus-response mappings and manual responding generalize to the Stroop and PWI tasks that we used, where stimulus-response mappings were not arbitrary and vocal responding was required.

One possible (though, not perfect) natural analog of the many-to-one mapping used by De Houwer (2003) and Van Veen and Carter (2005) is the picture categorization task (e.g. Glaser & Düngelhoff, 1984; Experiment 2). In this task, participants are instructed to name the category of the pictured object while ignoring the distractor word. For example, they say "animal" in response to a picture of a cat. The distractor word is the name of the picture (e.g., *cat*), comparable to the congruent condition of De Houwer. Alternatively, the distractor word is the name of an object from the same semantic category (e.g., *dog*, which is also linked to the response "animal"), comparable to the condition with a perceptual difference but the same response of De Houwer. Finally, the distractor word is the name of an object from another semantic category (e.g., *house*, which is linked to the response "building"), comparable to the condition with both a perceptual and response difference in the study of De Houwer. Glaser and Düngelhoff observed that RTs were longer when there was both a perceptual and response difference (i.e., a picture of a cat combined with the word *house*) than when there was a perceptual but not a response difference (i.e., a picture of a cat combined with the word *dog*), suggesting a locus of the effect in word planning. This is in line with the conclusions of De Houwer and Van Veen and Carter for the Stroop task with an arbitrary

stimulus-response mapping and manual responding. However, picture categorization RTs did not differ when there was no perceptual or response difference (i.e., a picture of a cat combined with the word *cat*) and when there is a perceptual but not a response difference (i.e., a picture of a cat combined with the word *dog*), suggesting that no effect arises in perceptual encoding, unlike what De Houwer and Van Veen and Carter concluded. This suggests that results obtained with arbitrary stimulus-response mappings and manual responding do not necessarily generalize to the Stroop and PWI tasks that we used, where stimulus-response mappings were not arbitrary and vocal responding was required. Thus, our conclusion that the Stroop-like effect arises during word planning in Stroop and PWI tasks is not at odds with the findings and conclusions of De Houwer and Van Veen and Carter.

To conclude, the present ERP study examined the Stroop-like effect in a single group of participants performing a standard color-word Stroop task (three colors and distractor words), a standard PWI task (39 pictures and distractor words), and a Stroop-like PWI task (three pictures and distractor words). In all three tasks, the Stroop-like effect was associated with a modulation of the N400, starting around 350 ms after stimulus onset. No earlier effects were found for the PWI tasks. The onset of the Stroop-like effect at about 350 ms in all three tasks links the effect to word planning rather than perceptual encoding. We conclude that the Stroop-like effect arises during word planning in both Stroop and PWI.

CHAPTER 3

CONTROL ADJUSTMENTS IN SPEAKING: ELECTROPHYSIOLOGY OF THE GRATTON EFFECT IN PICTURE NAMING

A slightly modified version of this chapter was published as:

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M., 2017 Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex*, *92*, 289-303.

ABSTRACT

Accumulating evidence suggests that spoken word production requires different amounts of top-down control depending on the prevailing circumstances. For example, during Stroop-like tasks, the interference in response time (RT) is typically larger following congruent trials than following incongruent trials. This effect is called the Gratton effect, and has been taken to reflect top-down control adjustments based on the previous trial type. Such control adjustments have been studied extensively in Stroop and Eriksen flanker tasks (mostly using manual responses), but not in the picture-word interference (PWI) task, which is a workhorse of language production research. In one of the few studies of the Gratton effect in PWI, Van Maanen and Van Rijn (2010) examined the effect in picture naming RTs during dual-task performance. Based on PWI effect differences between dual-task conditions, they argued that the functional locus of the PWI effect differs between post-congruent trials (i.e., locus in perceptual and conceptual encoding) and post-incongruent trials (i.e., locus in word planning). However, the dual-task procedure may have contaminated the results. We therefore performed an EEG study on the Gratton effect in a regular PWI task. We observed a PWI effect in the RTs, in the N400 component of the event-related brain potentials, and in the midfrontal theta power, regardless of the previous trial type. Moreover, the RTs, N400, and theta power reflected the Gratton effect. These results provide evidence that the PWI effect arises at the word planning stage following both congruent and incongruent trials, while the amount of top-down control changes depending on the previous trial type.

3.1. INTRODUCTION

Goal-oriented behavior relies on a flexible system of top-down control that allows for modification of processing strategies based on an analysis of costs and benefits of different processing types in order to improve performance on a task (e.g., Gratton, Coles, & Donchin, 1992). For example, in an experiment with congruent and incongruent stimuli (like in the color-word Stroop task with congruent and incongruent color-word combinations, such as the word *red* in green ink), a participant can adjust the extent to which distractor information is processed depending on the expected trial type (e.g., Lamers & Roelofs, 2011). In particular, participants may choose between a narrow or a wide strategy in stimulus processing depending on whether a more or less thorough analysis is expected to optimize performance. In an experiment, the participant may rely on the previous trial type to form an expectation about the next trial type (Gratton et al., 1992), expecting trial-type repetition (e.g., Egner, 2007). Alternatively, a participant's expectations can be determined by cues that designate the probability of the next trial type (e.g., Aarts, Roelofs, & Van Turennout, 2008; Aarts & Roelofs, 2011; Gratton et al., 1992) or by means of global probability of stimuli of different types (e.g., Carter, Macdonald, Botvinick, Ross, Stenger, Noll, & Cohen, 2000).

Control adjustments have been intensively studied using Stroop-like tasks. In a seminal study, Gratton et al. (1992) observed that in such tasks, the difference in response time (RT) between incongruent and congruent trials is typically larger following congruent trials than following incongruent trials, referred to as the *Gratton effect* in later research. The Gratton effect has been examined using Stroop and Eriksen flanker tasks (mostly employing manual responses), but not in the picture-word interference (PWI) task (e.g., Glaser & Düngelhoff, 1984), which has been much used in language production research. Recently, Van Maanen and Van Rijn (2010) examined the Gratton effect in PWI embedded in a dual-task paradigm (see below). Aside from reporting (expected) quantitative differences in RTs, they also argued that post-congruent and post-incongruent trials differ with respect to the functional locus of the PWI effect (i.e., the stage at which the incongruence of the picture and the distractor influences processing of the stimulus). We argue that the dual-task procedure might have contaminated the results. The aim of the research reported in the present chapter was to examine the Gratton effect in regular PWI and to use EEG to determine the functional locus of the PWI effect and its modulation by previous trial type.

In the following, we first describe the trial-to-trial manifestation of the Gratton effect in Stroop-like tasks in more detail. Next, we discuss the dual-task findings and locus-shift account of Van Maanen and Van Rijn (2010), and we present an alternative interpretation of their findings in terms of task scheduling rather than a shift in locus. Then, we discuss the results of previous EEG studies on the Gratton effect, which motivate an examination

of the Gratton effect in the N400 component of the event-related brain potentials and in frontal theta power. Next, the results of our EEG study are reported. Finally, we discuss the consequences of our findings for the debate about the functional locus of the PWI effect.

3.1.1. The Gratton effect in response times

Trial-to-trial sequential effects have been extensively studied using Stroop-like tasks, in which the participant is presented with stimuli that are combinations of a target dimension and a distractor dimension. The participant is instructed to respond to the target dimension while ignoring the distractor dimension. In a proportion of stimuli the target and the distractor dimensions activate the same response (i.e., congruent trials), while in the rest of stimuli the target and the distractor dimensions are associated with different responses (i.e., incongruent trials). For example, in the color-word Stroop task, the stimuli are color words that are printed in a certain color ink (e.g., the word *red* in red or green ink), and the participant is instructed to respond to the color while trying to ignore the word (MacLeod, 1991). In the Eriksen flanker task, the stimuli are strings of letters (e.g., HHSHH or SSSSS) and the participant is instructed to respond to the central (i.e., target) one, while ignoring the distracting flankers (i.e., Eriksen & Eriksen, 1974). The PWI task employs drawings of objects with superimposed object names, and the participant is instructed to name the picture while ignoring the word (Glaser & Düngelhoff, 1984). A common finding of Stroop-like tasks is that participants give slower and less accurate responses on incongruent trials as compared to congruent trials (Eriksen & Eriksen, 1974; Glaser & Düngelhoff, 1984; MacLeod, 1991). Critically, this Stroop-like effect is larger on trials following congruent trials than on trials following incongruent trials (e.g., Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Gratton et al., 1992; Lamers & Roelofs, 2011; Ullsperger, Bylsma, & Botvinick, 2005, for the Eriksen flanker task; Egner & Hirsch, 2005; Kerns, Cohen, MacDonald, Cho, Stenger, & Carter, 2004; Lamers & Roelofs, 2011; Notebaert, Gevers, Verbruggen, & Liefooghe, 2006, for the Stroop task). The Gratton effect has been taken to reflect adjustments in top-down control that a participant exerts in certain contexts (but see Hommel, Proctor, & Vu, 2004; Mayr, Awh, & Laurey, 2003).

The Gratton effect has been studied extensively in manual Stroop-like tasks (e.g., Botvinick et al., 1999; Egner & Hirsch, 2005; Kerns et al., 2004) and has been mostly interpreted using the conflict-monitoring account (Botvinick, Braver, Barch, Carter, & Cohen, 2001, but see Aarts et al., 2008; Lamers & Roelofs, 2011). This account suggests that the response conflict that occurs in incongruent trials leads to an increased top-down control during the following trial. The top-down control biases the processing of the stimulus by enhancing the target stimulus dimension or suppressing the irrelevant distractor dimension, thereby diminishing the magnitude of the Stroop-like effect.

Research on the Gratton effect in speech production paradigms has remained scarce (Lamers & Roelofs, 2011, for the Eriksen flanker task with vocal responding and the Stroop task; Freund, Gordon, & Nozari, 2016, and Van Maanen & Van Rijn, 2010, for the PWI task). Van Maanen and Van Rijn investigated the control adjustments using a dual-task procedure (Pashler, 1984). The participants had to perform two tasks at the same time: a tone classification task and a PWI task. In the PWI task, picture and distractor word were semantically related (e.g., a picture of a cat combined with the word *dog*, the incongruent condition), semantically unrelated (e.g., cat – *house*, the unrelated condition), or the distractor word was the name of the picture (cat – *cat*, the congruent condition). Critically, the interval between the tone and the PWI stimulus (stimulus-onset asynchrony, SOA) was manipulated. The SOAs were 100, 350, and 800 ms. The overall naming RTs decreased with increasing SOA. Moreover, Van Maanen and Van Rijn suggested that at the longest SOA (i.e., 800 ms) the semantic interference effect (i.e., the longer RTs observed for semantically related compared with unrelated trials) was smaller on trials following incongruent trials than following congruent trials, reflecting the Gratton effect. At the shorter SOAs, the semantic interference effect was present on trials following incongruent trials, but the semantic effect was absent on trials following congruent trials. Under the assumption of a response-selection bottleneck (Pashler, 1984), the absence of the semantic effect at short SOAs suggests that the effect arises before response selection in perceptual and conceptual encoding, whereas full-blown semantic interference suggests a locus in response selection or later (for an extensive discussion, see Piai, Roelofs, & Schriefers, 2014). Following congruent trials, the semantic effect was absent at short SOAs, suggesting a locus in perceptual and conceptual encoding, whereas following incongruent trials, the semantic effect was present, suggesting a locus in word planning or later. According to Van Maanen and Van Rijn, these findings indicate a shift in the functional locus of picture-word interference from perceptual and conceptual encoding on post-congruent trials to word planning on post-incongruent trials.

It is important to note that researchers have found no agreement on the functional locus of interference in the PWI task. Picture naming is generally assumed to consist of three main stages (e.g., Levelt, Roelofs, & Meyer, 1999; Roelofs, 2003, 2014). First, the picture is perceived and conceptually identified based on the stimulus features, hereafter the *perceptual and conceptual encoding* stage. According to a meta-analysis of Indefrey and Levelt (2004; Indefrey, 2011), this stage is completed around 200 ms post stimulus-onset. Then, the corresponding spoken word is planned based on information retrieved from long-term memory, hereafter the *word planning* stage. This stage starts about 200 ms post stimulus-onset and ends around 150 ms before articulation onset. Finally, the planned picture name is articulated. As we discussed in detail elsewhere (Shitova, Roelofs, Schriefers, Bastiaansen, & Schoffelen, 2016), the locus of the interference effect has been linked either to perceptual and conceptual encoding (Dell'Acqua, Job, Peressotti, & Pascali, 2007; Dell'Acqua, Sessa,

Peressotti, Mulatti, Navarrete, & Grainger, 2010; Van Maanen, Van Rijn, & Borst, 2009) or to word planning (Piai, Roelofs, & Schriefers, 2014; Schnur & Martin, 2012). The dual-task findings by Van Maanen and Van Rijn (2010) would suggest that the locus is in perceptual and conceptual encoding or word planning depending on the previous trial type.

Van Maanen and Van Rijn (2010) demonstrated the utility of their locus-shift account in computer simulations using the RACE/A model. The Gratton effect was instantiated in the model as a modulation of the speed of processing of words relative to pictures. Assuming that the participants exert more top-down control (i.e., "more suppression of the reading response", p. 175), on post-incongruent trials than on post-congruent trials, the speed of word processing was set lower for post-incongruent than for post-congruent trials in the computer simulations. As a consequence, the locus of the semantic effect was in perceptual and conceptual encoding for post-congruent trials but in word planning for post-incongruent trials in the model.

However, there is an alternative interpretation of the empirical results reported in the Van Maanen and Van Rijn (2010) study, which does not assume that the locus of the interference effect shifts depending on the previous trial type. Rather, the alternative interpretation holds that there is a single fixed locus of semantic and Stroop-like effects (i.e., in word planning, see Piai, Roelofs, & Schriefers, 2014; Shitova et al., 2016). The behavioral effects reported by Van Maanen and Van Rijn may also have resulted from a change in strategic task scheduling depending on the previous trial type, in order to optimally perform in the dual-task setting. Piai, Roelofs, Schriefers (2014) argued that participants may strategically control the amount of overlap between response-selection processes in the two tasks during dual-task performance. On post-congruent trials, the participants of Van Maanen and Van Rijn might have adopted a more daring strategy, allowing overlap in response-selection processes between the tone and PWI tasks. In this way, word planning for the PWI task could have started before the response for the tone classification task was selected, such that the semantic PWI effect overlaid the SOA effect and these effects were underadditive on post-congruent trials. However, on post-incongruent trials, the participants might have adopted a more conservative strategy, not allowing such overlap, i.e. word planning for the PWI task only started after the response for the tone classification task was selected, which made the SOA effect and the semantic effect additive on post-incongruent trials. Thus, this difference in task scheduling can explain the additivity of the SOA and the semantic effects on post-incongruent trials and the underadditivity on post-congruent trials, which was observed by Van Maanen and Van Rijn. Thus, the difference in semantic effect at short SOAs between post-congruent and post-incongruent trials does not need to reflect a shift in locus of the PWI effect, as Van Maanen and Van Rijn maintain, but may instead reflect a difference in strategic task scheduling.

If instead the previous trial type leads to a shift of the functional locus, as suggested by Van Maanen and Van Rijn (2010), this may have important implications for the interpretation of previous studies examining the functional locus of PWI effects. In such studies, congruent and incongruent trials are presented to participants in pseudo-randomized order and the critical measures are usually obtained through averaging trials within conditions irrespective of the previous trial type. However, if the locus of interference systematically shifts following congruent and incongruent trials the results of such averaging depend on the proportions of the transitions of each type, which are not routinely controlled. Thus, the question of whether the locus of interference in the PWI task indeed depends on the previous trial type is important, yet it needs to be investigated without a potentially confounding dual-task paradigm, which may contaminate the results as we indicated. Therefore, in the present study we investigated the behavioral and electrophysiological correlates of the Gratton effect in a regular picture-word interference task.

3.1.2. The Gratton effect in electrophysiological measures

Electrophysiological correlates of control adjustment effects have been observed in a number of Stroop-like tasks. A typical finding for the Eriksen flanker task is a larger N2 component (i.e., "second negative peak in the averaged ERP waveform," Folstein & Van Petten, 2008) over fronto-central electrode sites on incongruent compared to congruent trials (Kopp, Rist, & Mattler, 1996). This N2 effect was consistently smaller following incongruent than following congruent trials (Clayson & Larson, 2011; Larson, Clayson, & Clawson, 2014, for a review). For the Stroop task (in various adaptations), a prominent negative-going deflection at approximately 400-450 ms post-stimulus onset, an N400 component, is typically larger on incongruent than on congruent trials (Liotti, Woldorff, Perez III, & Mayberg, 2000; Hanslmayr, Pastötter, Bäuml, Gruber, Wimber, & Klimesch, 2008). Some of these previous studies referred to this component as an "N450", but we take it to be an N400 (see Piai, Roelofs, & Van der Meij, 2012; Piai, Roelofs, Jensen, Schoffelen, & Bonnefond, 2014; Roelofs, Piai, Garrido Rodriguez, & Chwilla, 2016; Shitova et al., 2016). Contrary to the N2 effect in the Eriksen flanker task, the N400 effect has not been consistently shown to follow the Gratton pattern of control adjustments in the Stroop task. Null results were reported for a manual color-word Stroop task (Larson, Kaufman, & Perlstein, 2009; Larson, Clawson, Clayson, & South, 2012) and an auditory Stroop task with manual, overt vocal, and covert vocal responses (Donohue, Liotti, Perez III, & Woldorff, 2011). West, Bailey, Tiernan, Boonsuk, and Gilbert (2012), however, reported that the N400 effect was larger following congruent trials than following incongruent trials in a counting Stroop task (for similar effects of global probability on control adjustments in the counting Stroop task, see West & Bailey, 2012).

Moreover, Larson, Clayson, Kirwan, and Weissman (2016) reported control adjustment effects on the N400 component in a novel prime-probe Stroop-like task. The N400 effect has also been reported to reflect the semantic and Stroop-like effects in the PWI task (e.g., Piai, Roelofs, & Van der Meij, 2012; Piai, Roelofs, Jensen, et al., 2014; Shitova et al., 2016). Based on the N400 findings, it has been argued that the semantic and Stroop-like effects in the PWI task arise during word planning rather than during perceptual and conceptual encoding. As indicated before, the perceptual and conceptual encoding stage is completed around 200 ms post stimulus-onset, which is well before the moment that the Stroop-like effects occurs in the ERPs. For example, in the study of Shitova et al. (2016), the onset of the Stroop-like effect was around 350 ms after stimulus onset. This indicates that the effect occurred during word planning, which starts 200 ms post stimulus-onset. However, the N400 effect has never been used for studying control adjustments in speaking.

As concerns induced patterns of electrophysiological activity, power changes in the theta-band (4 – 8 Hz) from around 400 ms poststimulus until response onset over medial frontal cortex have consistently been reported for Stroop-like tasks (e.g., Hanslmayr et al., 2008; Nigbur, Ivanova, & Stürmer, 2011; Pastötter, Dreisbach, & Bäuml, 2013) as well as for feedback and error processing (Cavanagh, Cohen, & Allen, 2009; Cavanagh, Zambrano-Vazquez, & Allen, 2012). Interestingly, Hanslmayr et al. showed for the manual Stroop task that theta-band power over the anterior cingulate cortex increased monotonically from trials that yielded the least conflict (i.e., congruent) to trials that yielded the most conflict (i.e., incongruent trials for which the target dimension matched the distractor dimension on the previous trial). Thus, midfrontal theta power reflected not only the presence of conflict, but also its degree, which makes this frequency band interesting for research on control adjustments. Midfrontal theta power effects have only been reported once for PWI: Piai, Roelofs, Jensen, et al. (2014) demonstrated that theta-band power over left superior frontal gyrus was larger on incongruent than congruent trials in the PWI task, reflecting competitive selection of words in picture naming.

To summarize, researchers have found no agreement on the functional locus of the Gratton effect in picture naming: perceptual and conceptual encoding versus word planning. However, only a few studies (Van Maanen & Van Rijn, 2010; Freund et al., 2016) have examined the Gratton effect in PWI. Therefore, we conducted a new study examining the Gratton effect in standard PWI and measured ERPs and oscillatory power to obtain evidence on the functional locus of the effect. In their meta-analysis, Indefrey and Levelt (2004; Indefrey, 2011) estimated that perceptual and conceptual encoding in picture naming is completed at about 200 ms after picture onset, after which word planning starts. If the hypothesis of Van Maanen and Van Rijn of a differential locus is correct, the Stroop-like effect in PWI should appear in different time windows depending on the previous trial type, i.e., during perceptual and conceptual

encoding on post-congruent trials (i.e., before 200 ms post-stimulus onset) or during word planning on post-incongruent trials (i.e., after 200 ms post-stimulus onset). However, if the functional locus of the Stroop-like effect does not depend on the previous trial type, it is expected to appear during word planning on both post-congruent and post-incongruent trials.

3.2. METHOD

3.2.1. Ethics Statement

The line of research within which the study was performed was approved by the local Ethics Committee (ECG-2013-2504-095-102) and followed the World Medical Association Declaration of Helsinki. An informed consent form was obtained from all participants before the experiment.

3.2.2. Participants

Twenty-five participants (19 to 30 years old, mean 22.3 years, ten male) took part in the study for course credits or monetary compensation. All participants were right-handed, native speakers of Dutch, with normal or corrected to normal vision, and reported no history of neurological disorders.

3.2.3. Stimuli and procedure

The participants were seated approximately 70 cm away from a computer screen. The participants were instructed to name the picture as fast and accurately as possible, while ignoring the distractor word. They were also informed that the stimuli would only appear in the center of the screen. They first familiarized themselves with the stimulus set using a print-out of all pictures with names used in the experiment. Then the participants performed a practice session of ten trials. Up to four additional practice blocks of 10 trials each were performed by the participants if they requested more practice or if the experimenter was not satisfied with the performance on the previous block(s). Following the training, the participants performed five blocks of the task (156 trials per block). Small breaks (up to 5 minutes) were offered to the participants between blocks. The PWI stimuli consisted of a line drawing (4×4 cm) with a superimposed distractor word printed in lowercase Arial 24pt letters (1.8 – 3.2 cm wide, 0.5 – 0.7 cm high).

A trial is described in Figure 3.1. It started with the presentation of four asterisks during 1200 ms during which the participants were instructed to blink. After that a blank screen appeared for a random time between 700 to 1000 ms. The stimulus appeared immediately after that and stayed on the screen for 700 ms, after which the screen remained blank for 800 ms. Trials with response latencies shorter than 300 ms or longer than 1500 ms were excluded from the analysis. An experimental session including EEG application and instruction took on average a little more than two hours, while the task alone took approximately 50 minutes. The experiment was programmed using Presentation software (NeuroBehavioral Systems).

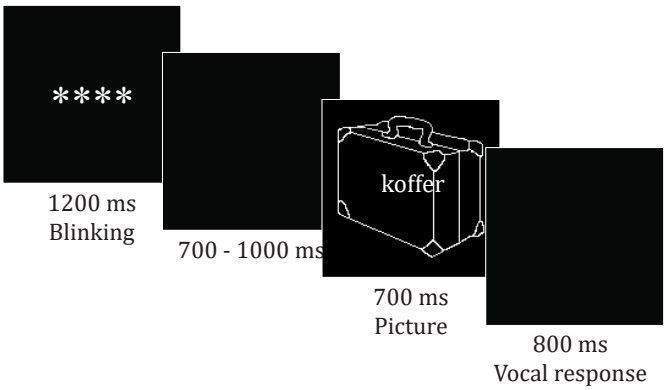

FIGURE 3.1. Trial events

The stimulus set was composed of 39 objects (13 groups of three items each). Pictures of objects were paired with distractor words to create congruent and incongruent stimuli: For a congruent stimulus a picture of an object was paired with its name, and for an incongruent stimulus a picture of an object was paired with the name of one of the other two objects from the same semantic domain (see the Appendix B1 for stimulus materials). The pictures were line drawings from a picture database of the Max Planck Institute for Psycholinguistics, from open source internet resources, or drawn from scratch. The pictures were previously used in a few naming experiments at our lab and showed high name agreement. The picture-word pairings were also used in our previous PWI study (Shitova et al., 2016) and showed a highly reliable Stroop-like effect.

A special pseudo-randomization technique was employed in order to create individual stimulus lists for each participant. The stimuli were pseudo-randomized using Mix software (Van Casteren & Davis, 2006) such that two stimuli that belonged to the same semantic group were separated by at least two other stimuli, the condition of the stimuli could not repeat

more than three times in a row, and the first letter of two neighboring stimuli could not be the same. The maximal difference between the total number of trials of each of the four types, i.e., congruent-after-congruent, congruent-after-incongruent, incongruent-after-congruent, and incongruent-after-incongruent (for each participant), was less or equal to 8. The participants were presented with the following number of trials per type (average ± standard error of the mean): congruent-after-congruent – 193.4±0.5; incongruent-after-congruent – 194.1±0.5; congruent-after-incongruent - 194±0.4; incongruent-after-incongruent – 193.4±0.5.

3.2.4. Recording

The participant's responses were recorded for offline semi-automatic response time estimation using PRAAT (Boersma, 2002). Naming errors were marked by the experimenter during the task. After per-participant and per-condition proportions of errors were calculated, error trials and post-error trials were discarded from analysis, both behavioral and EEG.

A standard 10-20 64-channel ActiCAP system (Brain Products) was used for scalp EEG measurements. The data were referenced online to FCz and rereferenced offline to the average of left and right mastoids. Four additional pairs of bipolar electrodes registered eye and lip movements, as well as the electrocardiogram (ECG). For the vertical electrooculogram (EOG) two electrodes were placed above and below the left eye. For the horizontal EOG two electrodes were placed on the left and right temples. For the lip electromyogram (EMG) two electrodes were placed at the left orbicularis oris superior and the right orbicularis oris inferior. For the ECG two electrodes were placed on the left and right collar bones. The ECG data were not used for the analysis reported in this chapter. The signals were digitized at a sampling rate of 1000 Hz and online-filtered with the low cutoff of 0.016 Hz and the high cutoff of 125 Hz.

3.2.5. Analysis

The data were analyzed with Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom analysis scripts using Matlab v.8.1.0.604 (R2013a, The MathWorks, Inc.). Epochs started at 500 ms before the onset of the stimulus and ended at the overt articulation onset (so, the length of trials was variable). The data were further inspected for EOG artifacts (blinks, eye movements) and artifactual trials were removed from the analysis. Then we inspected the data for muscular artifacts associated with speech onset in a manual trial-by-trial procedure. Such artifacts typically manifested as high-frequency noise over a spatially widespread set of EEG channels, starting at about 100 – 200 ms before the speech onset. If such artifact appeared earlier than 300 ms after the stimulus onset the trial was discarded completely, if the

artifact occurred later, only the contaminated part of the trial was removed. As a result of this procedure the EEG trials were of different lengths. All other types of artifacts including body movements and facial muscle artifacts not associated with speech were also discarded from the analysis via visual inspection.

Due to blinks, a median of 3.6% of the data were rejected (Q1 = 1.9%, Q3 = 5.7%). Through visual inspection, a median of 0.38% of the data were rejected (Q1 = 0, Q3 = 1.9%). Additionally, due to errors in naming (both on the current trial and on the previous trial), a median of 5.3% of the data were rejected (Q1 = 3.8%, Q3 = 8.0%). In total, a median of 10.5% of the data were excluded from analysis (Q1 = 6.8%, Q3 = 16.1%).

The data of one participant were discarded from the analysis due to extremely poor performance: The RTs were more than 2.5 standard deviations longer than the grand average RTs. The data of one more participant were discarded due to excessive blinking rate (almost 20% of the data) and residual eye movements that contaminated the rest of the data. This resulted in using the data of 23 participants for all types of analysis. Furthermore, electrode sites T7 and T8 were discarded from analysis due to high levels of noise in almost all participants' data.

For the ERP analysis the data were further band-pass filtered at 0.5-40 Hz with a onepass-zerophase linear non-causal hamming-windowed FIR filter and baseline-corrected using the last 300 ms of the pre-stimulus interval. The ERPs were computed through averaging epochs per condition and per participant. We created 8 condition-specific sets of ERPs per participant: current-congruent, current-incongruent, previous-congruent, previous-incongruent, congruent-after-congruent, congruent-after-incongruent, incongruent-after-congruent, incongruent-after-incongruent. Due to the difference in length between trials, the individual ERPs for the different conditions and participants also differed in length. Moreover, such individual ERPs were composed by diminishing number of trials as a function of time (since the response time differed across trials, see Figure B1).

In order to obtain the time-frequency representation (TFR) of the data, a sliding window Fourier Transform was used. We computed time-resolved power for frequencies within the 1 – 40 Hz band, using the entire time-window from 700 ms before stimulus onset until the end of the trial in steps of 10 ms. A variable length Hanning-tapered window was applied to estimate the power at each frequency using 3 oscillation cycles. Single-trial TFRs were averaged condition-specifically for each participant analogous to the ERPs. These individual TFRs were further baseline-corrected using a pre-stimulus interval from 300 ms before stimulus onset until stimulus onset, averaged over trials of all conditions. For the baseline

correction we first calculated baseline activity for every frequency, and then performed a relative change correction, i.e., subtraction of baseline activity values from post-stimulus power estimates with further division by the same baseline values.

The Stroop-like effect in ERPs and TFRs was assessed through contrasting incongruent-after-congruent trials with congruent-after-congruent trials (i.e., a post-congruent Stroop-like effect), and incongruent-after-incongruent trials with congruent-after-incongruent (i.e., a post-incongruent Stroop-like effect). The Gratton effect in ERPs and TFRs was assessed through contrasting difference ERPs and TFRs corresponding to post-congruent and post-incongruent Stroop-like effects.

Statistical significance of the difference between conditions in ERPs and TFRs was assessed using a cluster-based permutation approach (Maris & Oostenveld, 2007). For every contrast, a paired-samples *t*-statistic was calculated per electrode and time-point. Subsequently, these spatio-temporal maps were thresholded at a nominal *p*-value of 0.05 (two-sided). Next, clusters were created from the thresholded *t*-maps, where adjacent electrode and time-points were clustered, and the cluster statistic was calculated as the sum of the *t*-values within the cluster. The maximum cluster statistic was then compared against a permutation distribution, which was constructed through random re-labelling of the conditions of the original condition-specific ERPs or TFRs for 1000 times and determining spatio-temporal clusters as described earlier. Under the null hypothesis of exchangeability of ERPs or TFRs between conditions, the percentage of permutations that returned a larger cluster statistic than the original cluster statistic served a cluster-based *p*-value. The critical alpha-level for a one-sided and a two-sided test were put at 0.05 and 0.025 respectively. In order to assess statistical significance of the Stroop-like effects in ERPs we applied a two-sided test over the entire time-line (i.e., starting 500 ms before the stimulus onset and ending at the last time-point when individual ERPs for all participants and both conditions were available). For the analysis of the Gratton effect within the N400 time-window, the time-line was limited to 450 – 500 ms post-stimulus onset and a one-sided test was applied, since the difference between incongruent and congruent trials was expected to be larger following congruent rather than incongruent trials. Positive-sided tests were applied to TFRs. For the analysis of the Stroop-like effects the theta power was averaged per participant per condition (4 – 6 Hz) within the time-window of 350 – 650 ms post-stimulus onset, and these estimated entered the cluster-based permutation analysis. The Gratton effect was tested within the time-window of 350 – 550 ms post-stimulus onset.

3.3. RESULTS

3.3.1. Behavioral data

The group average RTs and error rates are presented in Figure 3.2. The figure shows that in both RTs and error rates the Stroop-like effect was smaller following incongruent trials than following congruent trials. In the RTs, the interaction was driven by the current-congruent trials (i.e., longer RTs on the congruent-after-incongruent trials than on the congruent-after-congruent trials), whereas it was driven by the current-incongruent trials in the error rates (i.e., lower error rates on the incongruent-after-incongruent trials than on the incongruent-after-congruent trials).

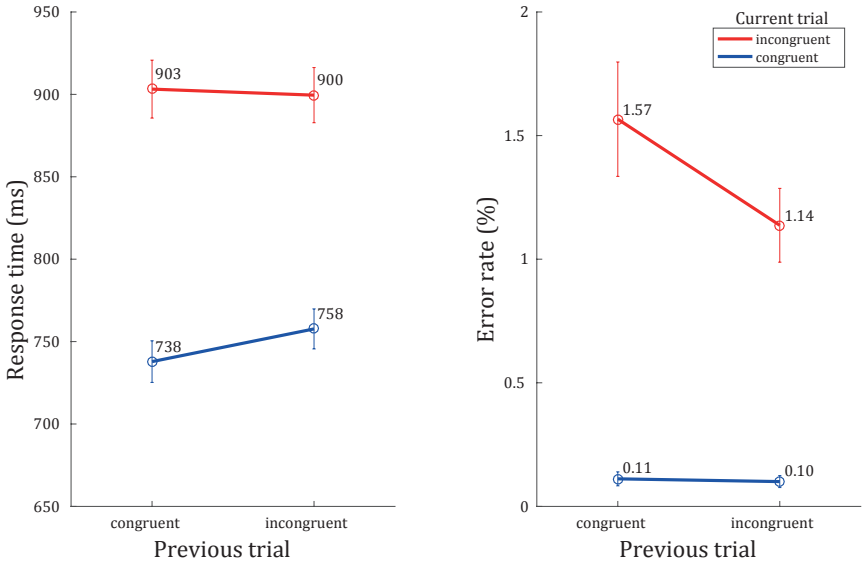

FIGURE 3.2. Mean response time and error rate as a function of previous trial (congruent, incongruent) and current trial (congruent, incongruent). The error bars indicate the standard error of the mean per condition.

For the RT data, a 2 × 2 repeated measures ANOVA with the factors *current trial* (congruent, incongruent) and *previous trial* (congruent, incongruent) showed significant main effects of *current trial* ($F(1,22) = 211.64$, $p < .001$), *previous trial* ($F(1,22) = 9.85$, $p = .005$), and an interaction between *current trial* and *previous trial* ($F(1,22) = 23.20$, $p < .001$). RTs in current-

congruent trials were slowed down if the previous trial was incongruent compared to congruent ($t(22) = 5.18$, $p < .001$). However, there was no influence of the previous trial type on RTs in current-incongruent trials ($t(22) = 1.13$, $p = .27$).

For the accuracy data, a 2 × 2 repeated measures ANOVA with the factors *current trial* (congruent, incongruent) and *previous trial* (congruent, incongruent) showed significant main effects of *current trial* ($F(1,22) = 48.41$, $p < .001$), *previous trial* ($F(1,22) = 16.24$, $p < .001$), and an interaction between *current trial* and *previous trial* ($F(1,22) = 10.38$, $p = .004$). A further pairwise *t*-test showed that error rates were reduced in current-incongruent trials if the previous trial was incongruent compared to congruent ($t(22) = 3.65$, $p = .001$). This effect was absent in current-congruent trials ($t(22) = 0.46$, $p = .64$). The interaction reflected that the Stroop-like effect (i.e., the difference between incongruent and congruent trials) was larger following congruent than following incongruent trials in both RTs and error rates, hence replicating the Gratton effect.

3.3.2. ERP data

Group-average ERPs on correct trials only for the four current-trial by previous-trial conditions over FCz are shown in Figure 3.3 (A; for group average ERPs over nine standard electrode sites see Appendix B2). Inspecting the curves visually, it can be observed that the waveforms deflected more negatively in the incongruent conditions than in the congruent conditions during the N400 time-window (approximately from 350 to 600 ms after the stimulus presentation). Moreover, as depicted in Figure 3.3B, the difference between the incongruent and congruent conditions (i.e. the Stroop-like effect) was larger following congruent than following incongruent trials, which is the Gratton effect.

A two-sided cluster-based permutation test performed on all time-points and all electrode sites yielded significant differences between congruent and incongruent stimuli (a Stroop-like effect) on both post-congruent and post-incongruent trials. Following congruent trials, the Stroop-like effect was present over a fronto-central cluster that was observed from approximately 360 until 760 ms post-stimulus onset ($p < .001$). Following incongruent trials, the Stroop-like effect emerged in a similarly distributed spatio-temporal cluster from approximately 400 until 770 ms post-stimulus onset ($p < .001$, see topographies of the Stroop-like N400 effects in Figure 3.3B).

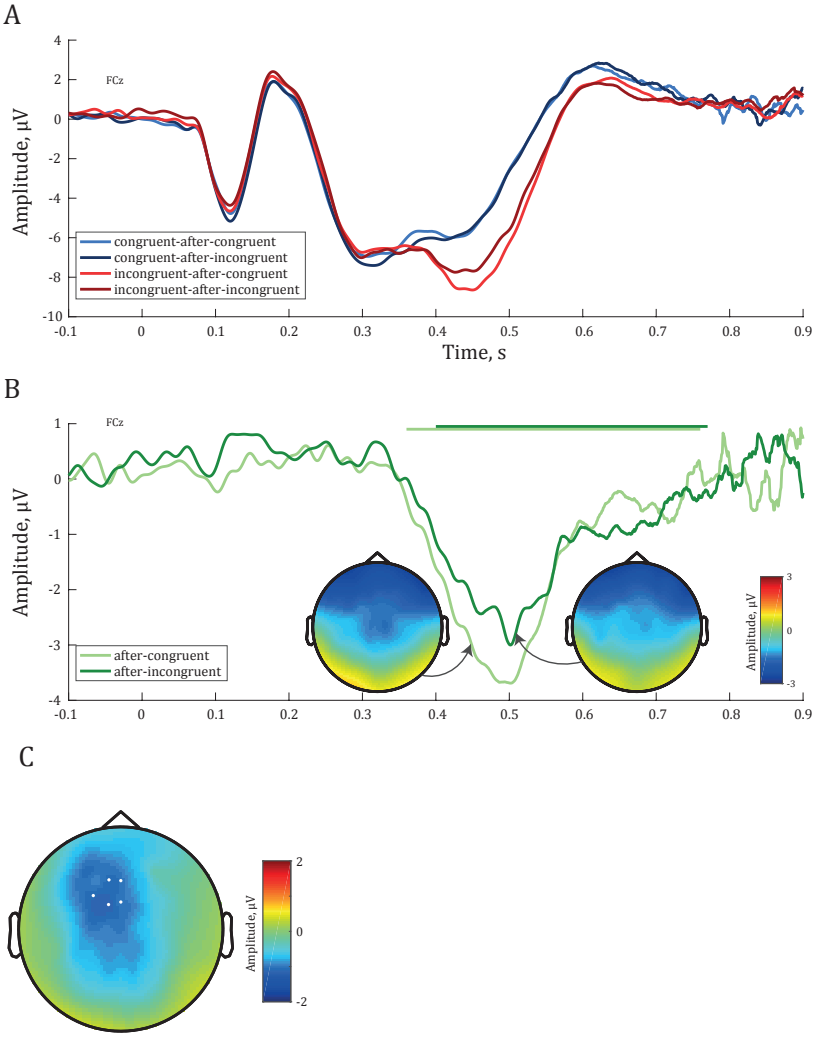

FIGURE 3.3. The N400 effects. **A.** Group-average sensor-level ERPs corresponding to the congruent-after-congruent, congruent-after-incongruent, incongruent-after-congruent, and incongruent-after-incongruent conditions, shown for FCz. **B.** Difference waveforms for after-congruent and after-incongruent trials, calculated as current-incongruent ERP minus current-congruent ERP. Horizontal lines mark the time-windows during which the Stroop-like N400 effects were significant in after-congruent and after-incongruent trials. Topographies of the Stroop-like effect in the N400 on after-congruent (left) and after-incongruent (right) trials, calculated through averaging amplitudes of the difference group ERPs within the time-windows in which the difference between current-congruent and current-incongruent conditions was significant. **C.** Topography of the Gratton effect, calculated through averaging amplitudes of the difference of the difference group ERPs within 450 - 500 ms post-stimulus onset. The highlighted electrode sites (F1, Fz, FC3, FC1, FCz) are those that entered the spatio-temporal cluster based on which we rejected the null hypothesis of equal magnitudes of the Stroop-like N400 effect in post-congruent and post-incongruent trials.

Visual inspection of difference waveforms suggested that the post-congruent and post-incongruent difference waveforms deflected differentially within the time interval of the N400 component (i.e., from approximately 350 to 600 ms post-stimulus onset) with the maximum difference reached within 450 - 500 ms post-stimulus onset over a range of midcentral electrode sites. Comparison of topographies of the Stroop-like effect after congruent and after incongruent trials revealed a difference across fronto-central electrode sites, slightly left-lateralized, centered around FC1 (see Figure 3.3C). A cluster-based permutation test showed a significant interaction between current-trial and previous-trial factors over the cluster of five fronto-central electrode sites (F1, Fz, FC3, FC1, FCz; p =.02): The negative-going difference between the current-incongruent and current-congruent trials within the 450 - 500 ms post-stimulus onset was larger following congruent trials than following incongruent trials, reflecting the Gratton effect in the N400. A post-hoc analysis of simple effects showed that within the time-window 450 – 500 ms post-stimulus onset the incongruent-after-congruent trials were associated with a larger negative deflection (i.e., the N400 component) than the incongruent-after-incongruent trials (p = .015), while no difference was observed between congruent-after-congruent and congruent-after-incongruent trials. Visual inspection of the distributions of peak latencies showed no systematic shift, and a two-sided t-test did not reveal difference between N400 peak latencies of the post-congruent and post-incongruent trials ($t(22)$ = -0.52, p = .61, CI = [-23.9, 14.4]). A two-sided Bayesian t-test confirmed absence of systematic difference in timing of the N400 (BF01 = 4.05, suggesting that the data were ~4 times more likely to be observed under the null hypothesis of no difference between conditions than under the alternative hypothesis of systematic shift in timing of the N400).

3.3.3. Time-frequency data

Figure 3.4 shows the time-frequency data (TFR). Visual inspection of group-averaged TFRs revealed theta-power (4-6 Hz) increase relative to the pre-stimulus baseline in all conditions, centered around midfrontal electrode sites (F1, Fz) and persisting from approximately 200 ms post-stimulus onset until the end of the trial.

Importantly, this event-related synchronization was larger on incongruent trials compared to congruent trials from approximately 350 to 650 ms post-stimulus onset (see Figure 3.4 for TFRs and topographies). This Stroop-like effect was significant in the cluster-based permutation test that included all electrode sites and all time-points between 350 and 650 ms post-stimulus onset (p = 0.03 for the post-congruent comparison; p = 0.04 for the post-incongruent comparison). Moreover, the Stroop-like effect in the midfrontal theta power was larger following congruent than incongruent trials between 350 and 550 ms post-stimulus onset, reflecting the Gratton effect in the theta power (p = 0.046).

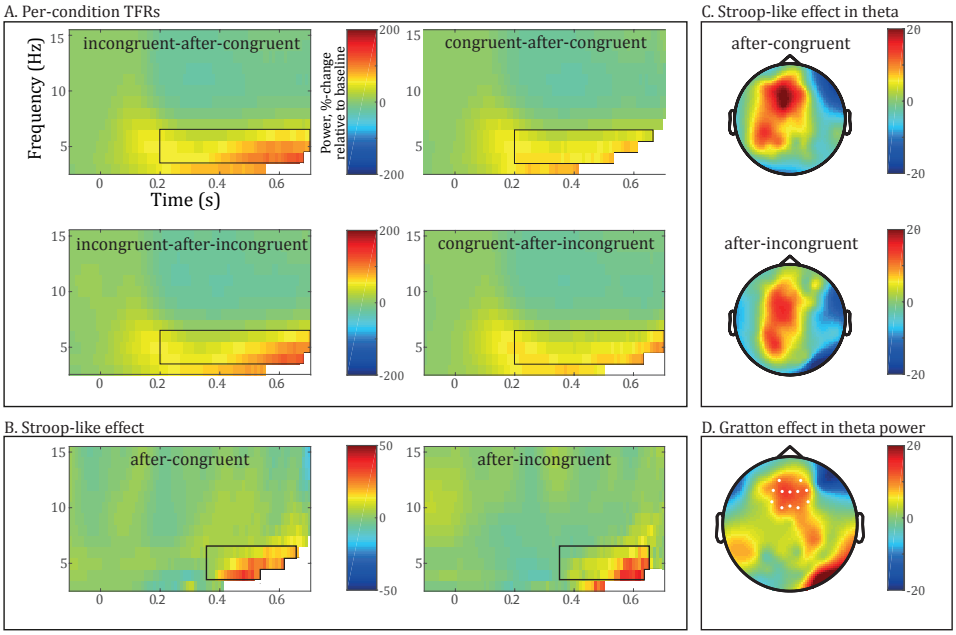

FIGURE 3.4. The theta-power effects. **A.** Group-average sensor-level TFRs corresponding to the incongruent-after-congruent, incongruent-after-incongruent, congruent-after-congruent, and congruent-after-incongruent conditions, averaged over 11 midcentral sites: AF3, F5, F3, F1, Fz, FC3, FC1, FCz, C3, C1, Cz. **B.** The Stroop-like theta power effect (i.e., the difference between the current-incongruent and current-congruent TFRs) for after-congruent (left) and after-incongruent (right) conditions. **C.** The Stroop-like theta power effect topographies for after-congruent (top) and after-incongruent (bottom) conditions. The topographical distributions were calculated through averaging power of the difference group TFRs between 350 and 650 ms post-stimulus onset. **D.** The Gratton theta-power effect topography. The topographical distribution of the Gratton effect was calculated through averaging power of the difference of the difference group TFRs between 350 and 550 ms post-stimulus onset. The highlighted electrode sites are those that entered the spatio-temporal cluster based on which we rejected the null hypothesis of equal magnitudes of the Stroop-like theta power effect in post-congruent and post-incongruent trials.

3.4. DISCUSSION

Researchers have found no agreement on the functional locus of the Gratton effect in picture naming: perceptual and conceptual encoding versus word planning. Above, we reported the results of an EEG study on the Gratton effect in regular PWI. Previous work indicates that the stage of perceptual and conceptual encoding is completed around 200 ms post stimulus-onset, followed by the stage of word planning that lasts until around 150 ms before articulation onset. We observed that the PWI effect was present in the RTs, N400, and frontal theta

power, regardless of previous trial type. Moreover, the RTs, N400, and theta power reflected the Gratton effect. The onset of the effects was about 350-400 ms post stimulus-onset, which is well after the completion of the stage of perceptual and conceptual encoding. Instead, the timing of the effects suggests that they arise during word planning. Below, we argue that these results challenge the locus-shift account of Van Maanen and Van Rijn. Our results indicate that the PWI effect arises at the word planning stage following both congruent and incongruent trials, while the amount of top-down control changes depending on the previous trial type.

3.4.1. Challenge to the locus-shift account

The only previous study that addressed the question of the locus of the Gratton effect in PWI (Van Maanen & Van Rijn, 2010) used a dual-task procedure. On long-SOA trials, the Gratton effect manifested itself as a numerical (but not statistically significant; see our re-analysis of their data below) reduction of the semantic effect on the trials following incongruent trials as compared to the trials following congruent trials. Furthermore, on short-SOA trials, the semantic effect vanished following congruent trials but remained following incongruent trials. Van Maanen and Van Rijn interpreted this pattern of results as evidence for a shift in the functional locus of the semantic interference effect: Early during processing (i.e., in perceptual and conceptual encoding) following congruent trials and late (i.e., in word planning) following incongruent trials. We argued that the pattern of short-SOA results by Van Maanen and Van Rijn could be explained differently, in terms of a task-scheduling difference between post-congruent and post-incongruent trials. Following congruent trials, the participants might have adopted a daring strategy that allowed overlap in response-selection processes related to the two tasks, which led to the absence of semantic interference at short SOAs. However, following incongruent trials, the participants might have used a conservative strategy that allowed little overlap, thus the semantic effect was present at short SOAs. Therefore, the difference in semantic effect at short SOAs between post-congruent and post-incongruent trials does not need to reflect a shift in locus of the PWI effect, as Van Maanen and Van Rijn maintain, but may instead reflect a difference in strategic task scheduling.

We therefore used a standard PWI task and recorded EEG, as well as behavioral data. We replicated the behavioral findings with respect to the Gratton effect (Freund et al., 2016): Both the RTs and the error rates showed that the Stroop-like effect diminished on the trials following incongruent trials as compared to the trials following congruent trials. As extensively discussed by Lamers and Roelofs (2011), studies in the literature show that the Gratton effect may be driven by current-incongruent trials (Kerns et al., 2004), current

congruent trials (Lamers & Roelofs, 2011), or both (Gratton et al., 1992). To explain this difference among studies in the underlying RT patterns giving rise to the Gratton effect (i.e., a larger Stroop-like effect on post-congruent than on post-incongruent trials), Lamers and Roelofs presented a two-factor account. According to this account, a difference in the extent to which distractor information is processed between post-congruent and post-incongruent trials gives rise to the basic Gratton effect, where a post-incongruent cost for congruent trials and a post-incongruent benefit for incongruent trials are observed. In addition, response caution is assumed to differ between post-congruent and post-incongruent trials. The caution is higher following post-incongruent trials equally prolonging RTs in the congruent and the incongruent condition on post-incongruent trials relative to post-congruent trials. In this way, the basic RT patterns underlying the Gratton effect are modulated, yielding the RT patterns observed by Kerns et al. (2004), Lamers and Roelofs (2011), and Gratton et al. (1992). In the present experiment, we observed the RT pattern obtained by Lamers and Roelofs. Greater caution on post-incongruent trials is likely to diminish error rates, therefore increasing the post-incongruent benefit for current-incongruent trials, as we saw in the present experiment. Lamers and Roelofs presented the results of computer simulations using the WEAVER++ model to show the utility of their two-factor account.

Moreover, we investigated the electrophysiological correlates of the Stroop-like effect (i.e., the N400 effect and the midfrontal theta power). The results paralleled the behavioral findings: The Stroop-like effect was present in the N400 component and the theta power regardless of the previous trial type, but it diminished after incongruent trials. No earlier effects were observed when comparing current congruent and current incongruent trials following either congruent or incongruent trials. Thus, the Gratton effect manifested itself only quantitatively (in RTs, N400 amplitude, and theta power) and we found no evidence for a shift of the locus of the Stroop-like effect depending on the previous trial type. Therefore, our findings challenge the conclusions of Van Maanen and Van Rijn.

Apart from design differences between the experiment by Van Maanen and Van Rijn (2010) and our experiment (i.e., a PWI task embedded in a dual-task paradigm vs. the standard PWI task), the studies employed different experimental contrasts to assess the Gratton effect in picture naming. In our study we focused on the Stroop-like effect (i.e., congruent vs. incongruent trials), while Van Maanen and Van Rijn studied the semantic effect (i.e., neutral vs. incongruent trials, also called unrelated vs. semantically related trials). Note that in the PWI task, the Stroop-like contrast includes the semantic contrast as a proper part. In a previous modeling study, Van Maanen and colleagues (2009) did not predict differences in the locus of the semantic and Stroop-like effects in PWI, and treated congruent and neutral trials as instances of a single control condition. Moreover, in previous electrophysiological studies (Piai, Roelofs, Jensen, et al., 2014; Roelofs et al., 2016), the Stroop-like effect and the semantic

effect were both reflected in the N400: There was a difference in N400 amplitude between incongruent and neutral trials (the semantic effect), and an even larger difference between congruent and incongruent trials (the Stroop-like effect). Thus, if the functional locus of the semantic effect is different between post-congruent and post-incongruent trials, as Van Maanen and Van Rijn claim, then this should also hold for the Stroop-like effect. Therefore, the Stroop-like effect should have been present much earlier in the ERPs at post-congruent trials (i.e., before 200 ms post picture-onset) than at post-incongruent trials (after 200 ms). In contrast, we observed that the Stroop-like effect was present in the N400 and had a similar onset for post-congruent and post-incongruent trials. Thus, our findings contradict the claim of Van Maanen and Van Rijn that the locus of the PWI effect differs between post-congruent and post-incongruent trials.

Van Maanen and Van Rijn (2010) suggested that they obtained a Gratton effect at the long SOA without reporting appropriate statistics. However, in a re-analysis of their data, we failed to obtain a Gratton effect at the long SOA for the semantic effect. *Current trial* and *previous trial* did not interact ($F(1,21) = 0.64$, $p = .43$), different from what is suggested by Van Maanen and Van Rijn (2010, p. 172). The Gratton effect was not present for the Stroop-like effect either ($F(1,21) = 0.1$, $p = .76$). These inconsistencies make the results of Van Maanen and Van Rijn problematic.

Since it has remained unclear how robust the findings of Van Maanen and Van Rijn (2010) are, we conducted additional analyses on the dual-task performance data collected by Piai, Roelofs, and Schriefers (2014, Experiment 2). The dual-task procedure used by Piai et al. was highly comparable to that of Van Maanen and Van Rijn. On every trial, a tone-classification stimulus was combined with a PWI stimulus in one of three conditions (i.e., congruent, incongruent, or neutral) either at zero-SOA (short-SOA condition) or at an SOA of 500 ms (long-SOA condition). We restricted our analysis to the semantic and Stroop-like effects in the critical short-SOA condition. For the semantic effect, a 2 × 2 ANOVA with factors *current trial* (neutral, incongruent) and *previous trial* (congruent, incongruent) showed a main effect of *current trial* ($F(1,20) = 16.0$, $p < .001$), but no main effect of *previous trial* ($F(1,20) = 0.14$, $p = .72$) or interaction of *current trial* and *previous trial* ($F(1,20) = 0.03$, $p = .87$). Similarly, for the Stroop-like effect, only a main effect of *current trial* ($F(1,20) = 13.61$, $p = .002$) was present, but no main effect of *previous trial* ($F(1,20) = 2.03$, $p = .17$) or interaction of *current trial* and *previous trial* ($F(1,20) = 2.19$, $p = .16$). The absence of an interaction of *current trial* and *previous trial* indicates that the magnitude of the semantic and the Stroop-like effects did not differ between post-congruent and post-incongruent trials at the short SOA, unlike what Van Maanen and Van Rijn observed for the semantic effect. Thus, in line with our present

electrophysiological findings, the dual-task data of Piai et al. do not provide evidence for a shift in the locus of the PWI effect depending on the previous trial type, which challenges the claim by Van Maanen and Van Rijn.

3.4.2. Electrophysiological manifestations of the Gratton effect

Besides presenting evidence against the idea of a shift in the locus of the Stroop-like effect in PWI, our study is the first to present evidence for a Gratton effect in the electrophysiological patterns of brain activity (i.e., the N400 component and theta power) using overt spoken responses. Several previous studies failed to observe control adjustment effects in the N400 amplitude using various adaptations of the Stroop task with manual or spoken responses (Donohue et al., 2011; Larson et al., 2009; Larson et al., 2012). Until now, only West and colleagues (2012) reported a Gratton-like modulation of the Stroop-like effect in a counting Stroop task with manual responses. Given that we did observe the Gratton effect using overt spoken responses, it is important to see in what way our study was different from previous EEG investigations in the field. First, one clear difference between PWI and Stroop is the stimulus set size. We used a relatively large stimulus set consisting of 39 pictures, while the classical Stroop task employs three or four colors. The large stimulus set is an obvious asset of PWI, because it prevents multiple repetitions of the same stimuli throughout the experiment, which has been shown to potentially invalidate top-down control interpretations of the Gratton effect (Hommel et al., 2004; Mayr et al., 2003; cf. Larson et al., 2016). Second, we used a larger overall number of trials in the experiment than some previous studies. This increased our statistical power and yielded a high signal-to-noise ratio (for a discussion of statistical power issues in control adjustments effects in Stroop-like tasks, see review Larson et al., 2014).

Interestingly, the way the Gratton effect manifested itself in our study was different for the RTs and the ERPs. In the RTs, the interaction between current trial type and previous trial type was driven by the current-congruent trials. RTs on congruent-after-incongruent trials were longer than RTs on congruent-after-congruent trials, while RTs on incongruent-after-congruent and incongruent-after-incongruent trials did not differ. However, the interaction in the N400 amplitude was driven by a stronger negative deflection on incongruent-after-congruent trials than on incongruent-after-incongruent trials, while the N400 amplitude on congruent-after-congruent and congruent-after-incongruent trials did not differ. As we discussed in the beginning of this section with respect to discrepancy between RT and error rates patterns, the Gratton effect may take different forms of interaction of current trial and previous trial (Lamers & Roelofs, 2011). The difference in

the way the Gratton effect is present in the RTs and ERPs in the present study suggests that an influence of response caution was present in the RTs, but not, or differently, in the ERPs. Greater caution after incongruent than after congruent trials also explains why error rates were reduced (see Figure 3.2). Future ERP studies may examine this account further.

To our knowledge, the present study is the first to directly investigate effects of control adjustments on midfrontal theta power in speech production. Midfrontal theta has been shown to originate from the medial frontal cortex including the anterior cingulate cortex (ACC) and it has been related to cognitive control, task switching, error detection, and feedback processing outside the language domain (Debener et al., 2005; Hanslmayr et al., 2008; Luu, Tucker, & Makeig, 2004). Until now only one study reported theta power effects in PWI (Piai, Roelofs, Jensen, et al., 2014). Piai et al. used magnetoencephalography (MEG). Therefore, a direct comparison of scalp-level topographies of the Stroop-like theta power effect observed in their study and in the present study is problematic. However, the topography observed in the current study is compatible with the left superior prefrontal source (presumably including portion of the ACC) of the theta activity estimated from the MEG data by Piai et al. Our findings regarding control adjustments in speaking revealed by midfrontal theta-band effects are also in line with previous studies investigating control adjustments in manual Stroop and Eriksen flanker tasks (Cavanagh et al., 2009; Hanslmayr et al., 2008; Nigbur et al., 2011).

It remains unclear to what extent our present N400 and theta power results represent the same or distinct brain activity. A comparison of the time-frequency representation of the ERP data and the standard single-trial time-frequency representation showed that the latter largely reflects phase-locked theta dynamics. Moreover, inter-trial coherence (ITC) analysis, which quantifies across trials the consistency of the frequency-specific phase, suggested a midfrontal theta effect over which the ITC values were increased on incongruent trials compared to congruent trials relative to other sites. However, in neither analysis did we observe a Gratton effect. A presence of the Gratton effect in the theta portion of the ERP and the ITC values would have suggested that our theta power and ERP effects are reflecting the same brain dynamics. However, given the absence of this effect, the current dataset does not allow to make claims regarding a same or distinct origin of the N400 and theta oscillations dynamics. Based on previous work in the field, it would be plausible that the theta activity observed in the current study was generated by the prefrontal cortex including ACC (e.g., Hanslmayr et al., 2008, Piai, Roelofs, Jensen, et al., 2014). However, the source of the N400 activity in PWI remains unclear, being either ACC (like in other Stroop-like tasks; Hanslmayr et al., 2008) or left middle temporal gyrus (Piai, Roelofs, Jensen, et al., 2014).

3.5. CONCLUSIONS

The present study provided evidence on control adjustments in speaking using the standard PWI task. The Gratton effect manifested itself in RTs, error rates, amplitude of the N400 component, and midfrontal theta power. The onset of the ERP effect at 350-400 ms post stimulus-onset (i.e., in the N400) suggests that the Stroop-like effect occurred at the word planning stage regardless of the previous trial type. This finding contradicts the account of Van Maanen and Van Rijn (2010) that posits that the functional locus of the PWI effect shifts depending on whether the previous trial is congruent or incongruent. Moreover, we observed a Gratton effect in the midfrontal theta power, which is in line with the research on the role of theta-band oscillations in cognitive control, until now almost exclusively reported for manual tasks. To conclude, our results provide evidence that the PWI effect arises at the word planning stage following both congruent and incongruent trials, while the amount of top-down control changes depending on the previous trial type.

CHAPTER 4

SPEAKING UNDER DISTRACTION: CONTROL ADJUSTMENTS IN WORD AND NOUN-PHRASE PRODUCTION

A slightly modified version of this chapter is under revision as:

Shitova, N. & Roelofs, A. Speaking under distraction: Control adjustments in word and noun-phrase production.

ABSTRACT

Control adjustments have been extensively studied using discrete responses in non-linguistic domains involving distraction. The standard behavioral pattern is a smaller difference in response time or error rate between congruent and incongruent trials (i.e., the Stroop-like effect) following incongruent trials than following congruent trials: the Gratton effect. The nature of control adjustments remains a question of debate: While some researchers maintain that the adjustments are driven by experienced response conflict (conflict-monitoring account), others believe that the adjustment effects are due to expected congruency. In a series of picture-word interference experiments, we obtained evidence that the control adjustments are driven by congruency rather than by conflict: The Stroop-like effect was larger on post-congruent trials than on post-neutral and post-incongruent trials, while there was no difference in the magnitude of the Stroop-like effect between post-incongruent and post-neutral trials. We observed the control adjustment effects not only in single-word but also in noun-phrase production, which demonstrates control adjustments in non-discrete response tasks. The implications of our findings for theories of performance monitoring in language production are discussed.

4.1. INTRODUCTION

Speakers can adjust the amount of top-down control they exert, depending on what strategy will lead to better performance while taking less effort (Duthoo, Abrahamse, Braem, Boehler, & Notebaert, 2014; Freund, Gordon, & Nozari, 2016; Lamers & Roelofs, 2011; Shitova, Roelofs, Schriefers, Bastiaansen, & Schoffelen, 2017). Exerting more control than needed (e.g., on a simple trial or task) is inefficient, whereas exerting less control than needed (e.g., on a difficult trial or task) is daring and may cause decrease in accuracy. Therefore, successful performance relies on flexible adjustment of strategies that the speaker employs. In certain experimental set-ups, participants get informed about the upcoming stimulus type through cues (e.g., Aarts & Roelofs, 2011; Aarts, Roelofs, & Van Turennout, 2008), through a predictive sequence of trial types (e.g., Rogers & Monsell 1995), or using probability distributions of trial types (e.g., Carter, Macdonald, Botvinick, Ross, Stenger, Noll, & Cohen, 2000). In other cases, participants can only rely on the trials to which they were previously exposed, expecting the same trial type to appear next (Egner, 2007).

In Stroop-like tasks, adjusting amounts of control depending on the previous trial type gives rise to the Gratton effect: The Stroop-like effect (i.e., difference in response time or error rate between congruent and incongruent trials) is larger following congruent trials than following incongruent trials (Gratton, Coles, & Donchin, 1992). For example, in the picture-word interference task (PWI), the difference in picture naming response time (RT) between congruent trials (e.g., a picture of a dog with the distractor word *dog* superimposed on it) and incongruent trials (e.g., a picture of a dog with the distractor word *cat* superimposed on it) is larger following a congruent trial than following an incongruent trial. The Gratton effect has been studied in spoken language production using PWI (Freund et al., 2016; Shitova, Roelofs, Schriefers, et al., 2017; Van Maanen & Van Rijn, 2010) as well as using the color-word Stroop and Eriksen flanker tasks (Lamers & Roelofs, 2011).

In spite of a large body of research existing on the topic of control adjustments, researchers found no agreement on the exact cognitive processes underlying the Gratton effect. The dominant account to date (at least, outside the language domain) posits that the decrease in the magnitude of the Stroop-like effect on post-incongruent trials relative to post-congruent trials is due to additional top-down control exerted following the detected response conflict on the previous incongruent trial, called *the conflict-monitoring account* (Botvinick, Braver, Barch, Carter, & Cohen, 2001). This account was also adopted in a few studies on language production within the language domain (Duthoo et al., 2014; Freund et al., 2016; Van Maanen & Van Rijn, 2010).

An alternative explanation of the Gratton effect was developed by Lamers and Roelofs (2011). They argued that the multitude of previously observed behavioral patterns underlying control adjustments (i.e., the Gratton effect may be present because of an RT difference on current congruent trials, current incongruent trials, or both) can be better explained with *expectancy of congruency*. On this model (and in line with the original proposal by Gratton et al., 1992), following a congruent trial a participant widens the attentional window, allowing more detailed processing of the distractor word. This attentional width change underlies the interaction between current and previous trial type in both RTs and error rates. A wider attentional window is beneficial for performance on congruent trials (therefore, the performance is better on post-congruent congruent trials than on post-incongruent congruent trials), but it worsens performance on incongruent trials (performance is worse on post-congruent incongruent trials than on post-incongruent incongruent trials). Furthermore, post-incongruent trials may be approached with more caution than post-congruent trials, which will slow down responses on both post-incongruent congruent and incongruent trials. Lamers and Roelofs demonstrated by computer simulations that this expectancy/caution account can explain the multitude of RT patterns underlying the Gratton effect in the literature.

Although the conflict monitoring and the expectancy accounts interpret the Gratton effect differently (conflict vs. expectancy), they generate the same predictions in terms of the magnitude of the Stroop-like effect for behavioral experiments that employ congruent and incongruent conditions only. However, following Lamers and Roelofs (2011), including a neutral condition allows one to adjudicate between the two accounts. Such neutral trials (e.g., a picture of a dog with a meaningless sequence of letters like *pqx* superimposed on it) evoke no response conflict, since the non-word sequence does not activate a word that would compete for selection with the picture name. Moreover, the neutral trials will not cause widening or shrinking of the attentional window and will not induce extra caution. Lamers and Roelofs maintained that if in an experiment the Stroop-like effect is larger on post-congruent and post-neutral trials than on post-incongruent trials, this would be evidence for the conflict-monitoring account. In contrast, if the Stroop-like effect on post-congruent trials is larger than on both post-neutral and post-incongruent trials, this would be evidence for the expectancy account of control adjustments. They observed in a variety of experiments employing versions of the Stroop and Eriksen tasks the latter pattern of the Gratton effect, which demonstrates that the expectancy/caution model better accounts for the experimental data than the conflict-monitoring account. However, Lamers and Roelofs employed single word production and did not assess picture naming, and therefore it remains to be seen whether their findings can be generalized to the PWI task and tasks employing non-discrete responses, such as phrase production.

Adjudicating between the conflict-monitoring and the expectancy accounts in language production is theoretically important for current models of speech monitoring. Nozari, Dell, and Schwartz (2011) and Nozari and Novick (2017) recently proposed a conflict-based account of monitoring for speech errors, which is largely inspired by the conflict-monitoring account, and more specifically, its application to error detection in response-conflict tasks (Yeung, Botvinick, & Cohen, 2004). The model of Nozari and colleagues assumes conflict-based error monitoring in speech production as well as conflict-based control adjustments, as examined using PWI by Freund et al. (2016) within the framework of this model. However, Freund et al. only used congruent and incongruent trials, thus it remains unclear whether the adjustments were driven by conflict (as they maintained) or expectancy of congruency.

In sum, the question of the nature of the Gratton effect in speaking still remains a matter of debate. In the present study, we specifically focused on the effects of post-neutral trials in an attempt to adjudicate between the conflict-monitoring account and the expectancy account of control adjustments.

4.2. EXPERIMENT 1

Control adjustments in picture naming have been studied in only a small number of studies (Duthoo et al., 2014; Freund et al., 2016; Shitova, Roelofs, Schriefers, et al., 2017; Van Maanen & Van Rijn, 2010). Moreover, the magnitude of control adjustment effects substantially differs among these studies (e.g., the Gratton effect is 10 ms in Freund et al., 23 ms in Shitova et al., and absent in Van Maanen & Van Rijn, 2010, see Shitova et al., for discussion). In the first experiment, we decided to attempt to replicate the Gratton effect in PWI with congruent and incongruent conditions. Furthermore, we employed a phrase production task in order to contribute to the growing body of research on phrase production (e.g., Bürki & Laganaro, 2014; Bürki, Sadat, Dubarry, & Alario, 2016; Sikora, Roelofs, & Hermans, 2016; Sikora, Roelofs, Hermans, & Knoors, 2016; Shitova, Roelofs, Coughler, & Schriefers, 2017), as well as to determine whether the control adjustment findings in tasks with discrete responses can be generalized to tasks with non-discrete responses. Participants named the pictures producing determiner + noun phrases in Dutch (e.g., in response to a picture of a goat, they said "de geit", the goat).

4.2.1. Method

Ethics statement. This and all the following experiments described in this chapter were conducted within the line of research approved by the local Ethics Committee (ECG-2013-2504-095-102) and adhered to the World Medical Association Declaration of Helsinki. All

participants were tested at the Donders Centre for Cognition. They signed an informed consent form prior to the experiment, after which they had an opportunity to ask questions related to the procedure of the experiment and protocols of data management within the institution.

Participants. Twenty-five participants (mean age 23.4 years, 18 women) took part in this experiment for monetary compensation or course credits. All participants were native speakers of Dutch, who were not raised bilingually, they had normal or corrected-to-normal vision and reported no history of neurological disease.

Procedure. The participants were seated in a normally-lit sound-proof booth, approximately 50 cm away from a computer screen (Benq XL2420Z, 1920×1080). A microphone was positioned close to their face in a way that it did not block the screen. The participants were instructed to name pictures as fast and accurately as possible, while ignoring the distractor word superimposed on the pictures. The participants were specifically instructed to name pictures using definite determiners in correspondence with the picture name (e.g., "de geit", the goat). Then they were presented with all pictures to be used in the experiment together with the expected names and determiners in the form of a print-out. The participants then performed a practice block of 10 trials randomly selected from the entire list of stimuli, but so that both trial types – congruent and incongruent – were presented. Then the participants could request up to four more practice blocks to ensure good performance during the experimental session. The overall number of trials was 480, split in 2 blocks of 240 trials each, with a short self-paced break in between (approximately 3 min). A trial started with presentation of a fixation cross in the center of the screen for 400 ms, following which the screen remained blank for a random time between 100 and 400 ms. The stimulus appeared on the screen for 700 ms, after which the screen remained blank again for 700 ms. The participants could respond while the picture was presented on the screen or during the following blank period. The experimental session including instruction and practice took approximately 30 minutes.

Stimuli. The pictures were small white line-drawings (4×4 cm) presented on a black background with a white distractor word printed on top of the picture in the center. The stimulus set consisted of 36 basic objects in 9 groups of 4 objects. The objects in groups were chosen in such a way that two names had neuter grammatical gender and required the determiner "het" in Dutch and the other two names were non-neuter and required the determiner "de". Pictures were paired with distractors to construct congruent or incongruent stimuli. A congruent stimulus was a picture with its name superimposed. An incongruent stimulus was a picture with the other same-gender name from the same group superimposed. The distractors were printed in Arial capital letters, size 20. The entire stimulus list with translations is presented in Appendix Table C1.

Individual lists of stimuli were created for each participant using Mix (Van Casteren & Davis, 2006). In order to avoid random facilitation and interference effects by repetitions of items belonging to the same group, determiners, or initial letters of words, the following restrictions were applied: The items from the same group should be separated by at least two items from other groups; the determiners should not be repeated more than twice in a row; the same condition (congruent, incongruent) cannot be repeated more than three times; the picture names with an identical initial letter should be separated by at least two other items.

Analysis. Overt responses were recorded for offline RT measurement, which was performed automatically using PRAAT (Boersma, 2002) and further controlled and, if needed, adjusted manually based on per-trial oscillograms and time-frequency spectra. Raw RT data were preprocessed and explored using Matlab v.8.5.0.197613 (R2015a, The MathWorks, Inc.) and further analyzed using the LME4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R core team, 2015). Since preliminary explorative analysis of the distributions of RTs revealed substantial right-skewness, a natural logarithm was taken of the RT values in order to obtain normal distributions, which is better suited for linear mixed-effects (LME) analysis. Errors only constituted 4.3% of all trials, therefore no specific error analysis was performed. Error trials and post-error trials were removed from all further analyses.

4.2.2. Results and Discussion

Group-averaged RTs and ERs are displayed in Figure 4.1. Factors current-trial (congruent and incongruent), previous-trial (congruent and incongruent), block, and interaction between current-trial and previous-trial entered the linear mixed-effects model as fixed effects. We also included random intercepts for subjects and items, as well as random slopes by-subject and by-item for the effects of current-trial and previous-trial. The results of this analysis are presented in Table 4.1.

Adding effects of current-trial, previous-trial, block, and interaction between current-trial and previous-trial significantly improved the model fit, as indicated by the *p*-values, obtained using likelihood tests comparing models with and without each of the effects (added sequentially in this order): current-trial ($\chi^2 = 74.1$, $p < .001$), previous-trial ($\chi^2 = 40.5$, $p < .001$), block ($\chi^2 = 205$, $p < .001$), Current-trial × Previous-trial ($\chi^2 = 10.4$, $p = .001$). The Stroop-like effect was smaller on post-incongruent trials (120 ms) than on post-congruent trials (132 ms), indicating a Gratton effect (12 ms, ß = -.02, $p = .001$).

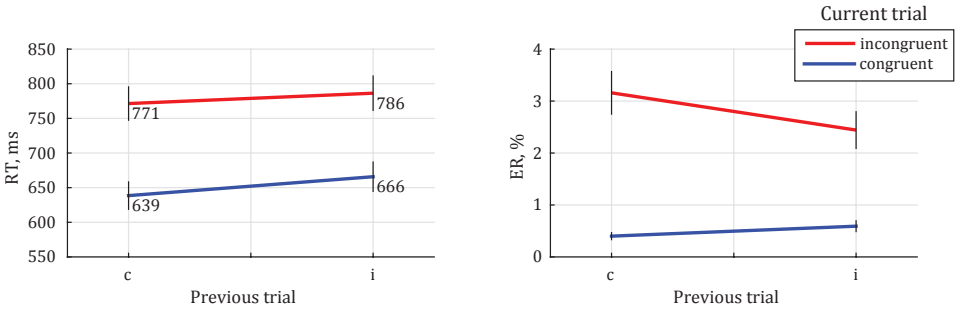

FIGURE 4.1. Group-averaged response time (RT) and error rate (ER) in Experiment 1 as a function of previous trial (congruent (c), and incongruent (i); exact RT values are specified next to the respective points on the graph). Error bars indicate the standard error of the mean per condition.

TABLE 4.1. Results of the LME analysis of Experiment 1

Fixed effect	Levels	*ß*	*SE*	*p*
Current-trial (relative to cong)	incongruent	.19	0.012	<.001
Previous-trial (relative to cong)	incongruent	.04	0.005	<.001
Block (relative to 1)	2	-.05	0.003	<.001
Current-trial × Previous-trial (relative to cur-cong prev-cong)	cur-incong prev-incong	-.02	0.006	.001

Note: cong = congruent, incong = incongruent, cur-cong = current-congruent, cur-incong = current-incongruent, prev-cong = previous-congruent, prev-incong = previous-incongruent

In the experiment, we replicated a standard control adjustment effect (i.e., the Gratton effect) in a noun-phrase production task with congruent and incongruent conditions. The materials and the experimental procedure were proven to be appropriate for studying the Gratton effect in picture naming.

4.3. EXPERIMENT 2

In the second experiment, we addressed the main question of the study: Are the control adjustments driven by previously experienced conflict, as suggested by the conflict-monitoring account, or by the expectancy of congruency? In order to do that, we included neutral trials in the stimulus list, besides congruent and incongruent stimuli. Comparing the magnitude of the Stroop-like effect on post-congruent, post-neutral, and post-incongruent trials allowed to adjudicate between the conflict-monitoring and the expectancy accounts.

4.3.1. Method

Twenty-eight volunteers (18 to 35 years old, 21 women) took part in the experiment for monetary compensation or course credits. One participant withdrew from the study after the first block due to extreme difficulty that he experienced when performing the task. The remaining 27 participants were 24.3 years old on average. The enrollment criteria for the participants, the procedure, the trial definition, and the stimuli were identical to those used in Experiment 1, except for the fact that neutral trials were added to the stimulus list. A neutral stimulus was a picture with a non-word sequence of consonants superimposed, such that the length of the non-word matched the length of the picture name in number of letters (see Appendix C1). The experimental session including instruction and practice took approximately 60 minutes.

4.3.2. Results and Discussion

Group-averaged RTs and ERs are displayed in Figure 4.2. A linear mixed-effects model was fitted using the same fixed and random effects as in Experiment 1. The results of this analysis are presented in Table 4.2.

Ln-transformed RTs were significantly influenced by all fixed factors: current-trial ($\chi^2 = 70.4$, $p < .001$), previous-trial ($\chi^2 = 32.1$, $p < .001$), block ($\chi^2 = 1352.5$, $p < .001$), Current-trial × Previous-trial ($\chi^2 = 16.6$, $p = .002$). As indicated by Table 4.2 and Figure 4.2, the Stroop-like effect on post-incongruent trials (117 ms) was smaller than on post-congruent trials (124 ms), indicating a Gratton effect (7 ms, ß = -.02, $p < .001$). Moreover, the Stroop-like effect on post-neutral trials (114 ms) was also smaller than on post-congruent trials (10 ms, ß = -.02, $p = .005$), suggesting that the Gratton effect was driven by expected congruency.

In the current phrase-production experiment we specifically studied whether control adjustments are driven by previously experienced conflict or whether they are driven by expectancy of congruency. The Stroop-like effect was larger on post-congruent trials than on post-neutral and post-incongruent trials. This finding is in line with the expectancy account, but it disagrees with the conflict-monitoring account. To make sure that this pattern of results is observed for noun phrases as well as single words, as used by Freund et al. (2016), we decided to run an experiment that employed single word responses, but which was otherwise identical to Experiment 2.

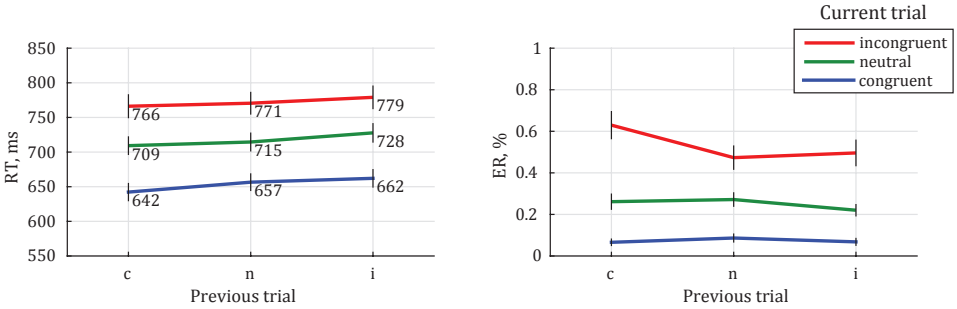

FIGURE 4.2. Group-averaged response time (RT) and error rate (ER) in Experiment 2 as a function of previous trial (congruent (c), neutral (n), and incongruent (i)). Error bars indicate the standard error of the mean per condition.

TABLE 4.2. Results of the LME analysis of Experiment 2

Fixed effect	Levels	*ß*	*SE*	*p*
Current-trial (relative to cong)	incong	.18	0.012	<.001
	neu	.10	0.009	<.001
Previous-trial (relative to cong)	incong	.03	0.004	<.001
	neu	.02	0.004	<.001
Block (relative to 1)	2	-.06	0.003	<.001
	3	-.07	0.003	<.001
	4	-.09	0.003	<.001
	5	-.11	0.003	<.001
Current-trial × Previous-trial (relative to cur-cong prev-cong)	cur-incong prev-incong	-.02	0.006	<.001
	cur-neu prev-incong	-.01	0.006	.10
	cur-incong prev-neu	-.02	0.006	.005
	cur-neu prev-neu	-.01	0.006	.02

Note: cong = congruent, incong = incongruent, neu = neutral, cur-cong = current-congruent, cur-incong = current-incongruent, cur-neu = current-neutral, prev-cong = previous-congruent, prev-incong = previous-incongruent, prev-neu = previous-neutral

4.4. EXPERIMENT 3

In the third experiment, we examined whether the pattern of effects observed in Experiment 2 for noun phrases is also observed for picture naming using single words. The experiment included neutral trials, as in Experiment 2, but required single word responses.

4.4.1. Method

Twenty-five participants (24.1 years old on average, 19 women) volunteered for the experiment. All enrolment criteria were identical to those described for Experiment 1. The procedure, including number of blocks and number of trials per block (1080 trials in 5 blocks) was identical to the procedure described for Experiment 2, with the only exception being the instruction that the participants received. Different from Experiment 2, where the participants were explicitly asked to name pictures using definite determiners, in Experiment 3 they were asked to name pictures using single words. The familiarization print-out was also adjusted so that the determiners did not appear next to the object names. Stimuli and analysis were identical to those described for Experiment 2. An experimental session took approximately 60 minutes.

4.4.2. Results and Discussion

Group-averaged RTs are shown in Figure 4.3. A linear mixed-effects model was fitted using the same fixed and random effects as in Experiments 1 and 2. The results of this analysis are presented in Table 4.3.

The fit of the model was significantly improved by including each of the fixed effects: current-trial (χ^2 = 74.3, p < .001), previous-trial (χ^2 = 19.2, p < .001), block (χ^2 = 871.1, p < .001), Current-trial × Previous-trial (χ^2 = 35.2, p < .001). As indicated by Table 4.3 and Figure 4.3, the Stroop-like effect was smaller on post-incongruent trials (135 ms) than on post-congruent trials (148 ms), indicating a Gratton effect (13 ms, ß = -.03, p < .001). Furthermore, the Stroop-like effect on post-neutral trials (134 ms) was smaller than on post-congruent trials (14 ms, ß = -.03, p < .001).

The current experiment was identical to Experiment 2 apart from the fact that the participants were not asked to use definite determiners but only nouns when naming pictures. Therefore, instead of testing noun-phrase production, this experiment probed single-word production. The results of Experiment 3 were similar to those reported for Experiment 2. There was a Gratton effect, i.e., the Stroop-like effect was larger on post-congruent trials than on post-incongruent trials. Moreover, the Stroop-like effect on post-congruent trials was larger than on post-neutral trials, replicating the pattern of results of Experiment 2.

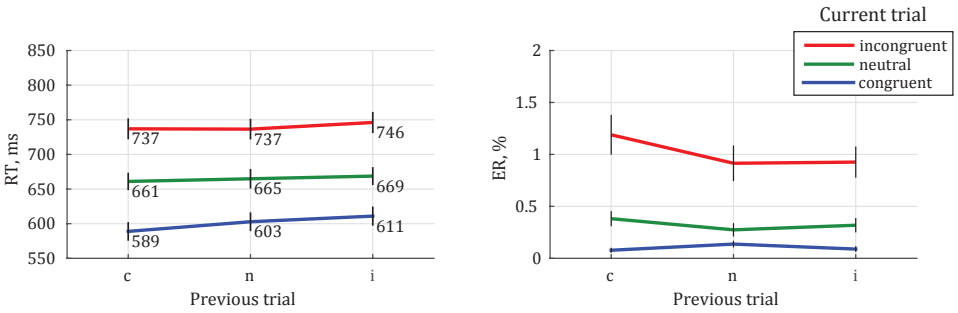

FIGURE 4.3. Group-averaged response time (RT) and error rate (ER) in Experiment 3 as a function of previous trial (congruent (c), neutral (n), and incongruent (i)). Error bars indicate the standard error of the mean per condition.

TABLE 4.3. Results of the LME analysis of Experiment 3

Fixed effect	Levels	*ß*	*SE*	*p*
Current-trial (relative to cong)	incong	.23	0.014	<.001
	neu	.12	0.010	<.001
Previous-trial (relative to cong)	incong	.04	0.005	<.001
	neu	.02	0.005	<.001
Block (relative to 1)	2	-.06	0.003	<.001
	3	-.08	0.003	<.001
	4	-.08	0.003	<.001
	5	-.08	0.003	<.001
Current-trial × Previous-trial (relative to cur-cong prev-cong)	cur-incong prev-incong	-.03	0.006	<.001
	cur-neu prev-incong	-.03	0.006	<.001
	cur-incong prev-neu	-.03	0.006	<.001
	cur-neu prev-neu	-.02	0.006	<.001

Note: cong = congruent, incong = incongruent, neu = neutral, cur-cong = current-congruent, cur-incong = current-incongruent, cur-neu = current-neutral, prev-cong = previous-congruent, prev-incong = previous-incongruent, prev-neu = previous-neutral

4.5. COMBINED ANALYSIS OF EXPERIMENTS 2 AND 3

In order to further test whether previous-incongruent or previous-congruent trials drive the Gratton effect in both single-word and noun-phrase production, we compared Stroop-like effects (current-congruent vs. current-incongruent) in post-congruent, post-neutral,

and post-incongruent trials. Given that Experiment 2 and Experiment 3 employed the same set of conditions (congruent, incongruent, and neutral) and same stimuli, we performed the analysis on the combined RT data of both experiments. A linear mixed-effects model included current-trial (congruent, incongruent, and neutral), previous-trial (congruent, incongruent, and neutral), block, and experiment, as well as interaction between current-trial and previous-trial as fixed effects. As random effects, intercepts for subjects and items, as well as random slopes by-subject and by-item for the effects of current-trial and previous-trial were included in the model.

This analysis replicated the results obtained for Experiment 2 and Experiment 3 separately. The current-trial, previous-trial, block, experiment, and interaction between current-trial and previous-trial had a significant effect on the ln-transformed RTs, based on the *p*-values, obtained using likelihood tests comparing models with and without each of the effects: current-trial (χ^2 = 97.9, p < .001), previous-trial (χ^2 = 51, p < .001), block (χ^2 = 2117, p < .001), experiment (χ^2 = 7.7, p = .006), Current-trial × Previous-trial (χ^2 = 46.6, p < .001). The results of this analysis are presented in Table 4.4. The Stroop-like effect was smaller on post-incongruent trials (126 ms, β = -.03, p < .001) and on post-neutral trials (124 ms, β = -.02, p < .001) than on post-congruent trials (136 ms).

TABLE 4.4. Results of the combined analysis of the data of Experiments 2 and 3

Fixed effect	**Levels**	***ß***	***SE***	***p***
Current-trial (relative to cong)	incong	.20	0.01	<.001
	neu	.11	0.01	<.001
Previous-trial (relative to cong)	incong	.03	0.003	<.001
	neu	.02	0.003	<.001
Block (relative to 1)	2	-.06	0.002	<.001
	3	-.07	0.002	<.001
	4	-.08	0.002	<.001
	5	-.09	0.002	<.001
Experiment (relative to Experiment 1)	Experiment 3	-.08	0.02	.002
Current-trial × Previous-trial (relative to cur-cong prev-cong)	cur-incong prev-incong	-.03	0.004	<.001
	cur-neu prev-incong	-.02	0.004	<.001
	cur-incong prev-neu	-.02	0.004	<.001
	cur-neu prev-neu	-.02	0.004	<.001

Note: cong = congruent, incong = incongruent, neu = neutral, cur-cong = current-congruent, cur-incong = current-incongruent, cur-neu = current-neutral, prev-cong = previous-congruent, prev-incong = previous-incongruent, prev-neu = previous-neutral

In order to address the main question of the current study – whether control adjustments are driven by experienced conflict (i.e., incongruent trials) or by expected congruency (i.e., congruent trials) – we reran the analysis after releveling the data, so that the effects of current-trial were calculated relative to the previous-neutral trials. This recomputed LME model revealed that the Stroop-like effect was larger on post-congruent trials than on post-neutral trials (ß = .02, *SE* = 0.004, $p < .001$), while the magnitude of the Stroop-like effect on post-incongruent trials did not differ from that on post-neutral trials (ß = .006, *SE* = 0.004, $p = .13$).

According to the conflict-monitoring account (Duthoo et al., 2014; Freund et al., 2016), control adjustments take place because of the response conflict that is detected on incongruent trials. Alternatively (Lamers & Roelofs, 2011), control adjustments are driven by expected congruent trials. The two accounts only make different predictions when neutral trials are employed in the task, which were not used in previous studies on the Gratton effect in picture naming (Duthoo et al., 2014; Freund et al., 2016; Shitova, Roelofs, Schriefers, et al., 2017). In the current analysis, we demonstrated that the Stroop-like effect (congruent vs. incongruent) was larger on post-congruent trials than on post-neutral trials. In contrast, incongruent trials do not additionally decrease the magnitude of the Stroop-like effect, compared to the neutral trials. This result suggests a post-congruent widening of the attentional window, which is beneficial for current-congruent trials. This finding is in line with the expectancy account of Lamers and Roelofs, and it challenges the conflict-monitoring accounted proposed by Duthoo et al. and Freund et al.

4.6. GENERAL DISCUSSION

We reported a series of experiments in which we investigated the nature of the Gratton effect in order to adjudicate between two accounts of control adjustments: conflict monitoring vs. expectancy of congruency. Moreover, we tested whether the pattern of control adjustments is also observed in non-discrete response tasks.

In Experiment 1, which involved a PWI task with phrase production in congruent and incongruent conditions, we replicated the Gratton effect in phrase production and confirmed that our stimuli and procedure could be used to further study the nature of control adjustments in language production. In Experiment 2, again using phrase production, we additionally introduced neutral trials and specifically addressed the question of whether the magnitude of the Stroop-like effect is different on post-congruent vs. post-incongruent trials and on post-congruent vs. post-neutral trials. We observed a larger Stroop-like effect on post-congruent trials than on post-incongruent and post-neutral trials, which agrees with the expectancy account and disagrees with the conflict-monitoring account. Experiment 3 was identical to

Experiment 2 except that the participants were now instructed to name pictures using single words. Experiment 3 fully replicated the pattern of results that was observed in Experiment 2, confirming that the pattern holds for single-word and phrase production. A combined analysis of the data of Experiments 2 and 3 showed that the Stroop-like effect was larger on post-congruent trials than on post-neutral and post-incongruent trials, and that there was no statistical difference in the magnitude of the Stroop-like effect between post-incongruent and post-neutral trials.

These results are in line with the expectancy account that predicts substantial facilitation on current-congruent previous-congruent trials compared to current-congruent previous-neutral and previous-incongruent trials, while the magnitude of the Stroop-like effect remains the same on post-neutral and post-incongruent trials. However, our results contradict the conflict-monitoring account that predicts a smaller Stroop-like effect on post-incongruent than on post-neutral and post-congruent trials. The results of the current series of experiments therefore challenge the general conflict-based speech-monitoring account proposed by Nozari et al. (2011).

To our knowledge, the current study is the first to provide evidence on control adjustments for non-discrete responses. The pattern of control adjustments appeared highly similar in Experiments 2 and 3 which only differed in the type of response the participants had to execute: a simple noun phrase vs. a single noun. The RTs in Experiment 2 (phrase production) were somewhat longer than those in Experiment 3 (single-word production; see results of the combined analysis), which was expected because producing a simple noun phrase involves retrieval of grammatical gender information as well as encoding of the determiner.

In all three experiments, we observed longer RTs on post-incongruent trials than on post-congruent trials. Such a pattern was previously observed in the PWI task (Duthoo et al., 2014) and was attributed by Lamers and Roelofs (2011) to different levels of caution that participants adopt following different types of trials. However, such post-conflict slowing (Verguts, Notebaert, Kunde, & Wühr, 2011) does not appear to impact control adjustment pattern consistently. For instance, we did not observe such an effect in our previous study on the Gratton effect (Shitova, Roelofs, Schriefers, et al., 2017), nor did many studies outside the language domain report post-conflict slowing (e.g., Kerns et al., 2004; Notebaert et al., 2006; Egner & Hirsch, 2005). It remains to be investigated which experimental manipulations trigger caution effects and how post-conflict slowing affects the general RT patterns underlying control adjustments.

4.7. CONCLUSION

In a series of experiments, we investigated control adjustments in single-word and noun-phrase production, specifically targeting the Gratton effect. We obtained evidence that the adjustments were driven by expected congruency rather than conflict, challenging the conflict-monitoring account of control adjustments in language production.

CHAPTER 5

P3 ERP COMPONENT REFLECTS ALLOCATION AND USE OF PROCESSING CAPACITY IN LANGUAGE PRODUCTION

A slightly modified version of this chapter was published as:

Shitova, N., Roelofs, A., Coughler, C., & Schriefers, H., 2017
P3 event-related brain potential reflects allocation and use of central processing capacity in language production. *Neuropsychologia, 106,* 138-145.

ABSTRACT

Allocation and use of central processing capacity have been associated with the P3 event-related brain potential amplitude in a large variety of non-linguistic tasks. However, little is known about the P3 in spoken language production. Moreover, the few studies that are available report opposing P3 effects when task complexity is manipulated. We investigated allocation and use of central processing capacity in a spoken phrase production task: Participants switched every second trial between describing pictures using noun phrases with one adjective (size only; simple condition, e.g., "the big desk") or two adjectives (size and color; complex condition, e.g., "the big red desk"). Capacity allocation was manipulated by complexity, and capacity use by switching. Response time (RT) was longer for complex than for simple trials. Moreover, complexity and switching interacted: RTs were longer on switch than on repeat trials for simple phrases but shorter on switch than on repeat trials for complex phrases. P3 amplitude increased with complexity. Moreover, complexity and switching interacted: The complexity effect was larger on the switch trials than on the repeat trials. These results provide evidence that the allocation and use of central processing capacity in language production are differentially reflected in the P3 amplitude.

5.1. INTRODUCTION

Language production involves planning and monitoring, which does not happen fully automatically but requires some central processing capacity (e.g., Roelofs & Piai, 2011, for a recent review). It is generally assumed that central processing capacity is limited and needs to be selectively allocated to ongoing mental processes. The amount of capacity that is allocated depends on an evaluation of the circumstances and task demands (e.g., Kahneman, 1973). In planning a word, phrase, or sentence, speakers need to engage in conceptual, syntactic, morphological, phonological, and phonetic encoding processes (e.g., Levelt, 1989). How much central processing capacity is allocated to some of these processes (e.g., conceptual encoding) may depend on the complexity of the intended utterance. For example, planning the production of "a cup" would seem to require less processing capacity than "a big cup", while planning to say "a big red cup" presumably requires even more capacity.

The allocation and use of central processing capacity have been studied extensively in non-linguistic domains by measuring event-related brain potentials (ERPs), thereby focusing on the P3 component (e.g., Kok, 2001; Polich, 2007; Polich & Kok, 1995). However, to our knowledge, there exist only three studies that investigated central processing capacity in language production using ERPs (Habets, Jansma, & Münte, 2008; Marek, Habets, Jansma, Nager, & Münte, 2007; Sikora, Roelofs, & Hermans, 2016). In all three studies, allocation and use of central processing capacity were taken to be modulated by phrase and sentence complexity, which was reflected in the P3 component. This corresponds to the effect of complexity on the P3 that has been observed for a variety of non-linguistic tasks (e.g., Brydges, Fox, Reid, & Anderson, 2014; Evans, Selinger, & Pollak, 2011; Kramer, Sirevaag, & Braune, 1987; Kok, 2001; Strayer & Drews, 2007; Watter, Geffen, & Geffen, 2001). However, the directionality of the P3 effect in language production differed between the studies of Habets et al. and Marek et al., on the one hand, and the study of Sikora et al., on the other.

In the ERP study of Habets et al. (2008), participants saw two sequentially presented pictures of objects (e.g., a car and a book), and they had to produce sentences describing two actions (e.g., driving and reading) involving these objects. A cue presented after the objects indicated whether the participants had to express the chronological order of object presentation (e.g., "After I drove the car, I read a book") or the non-chronological order (e.g., "Before I read a book, I drove the car"). It was assumed that participants would allocate more central capacity to the planning processes in the non-chronological than in the chronological condition. In line with this assumption, in the ERPs time-locked to the presentation of the cue, the amplitude of the P3 component was higher in the non-chronological (i.e., complex) than in the chronological (i.e., simple) condition.

A similar effect on the amplitude of the P3 was observed by Marek et al. (2007). In their study, the participants were shown networks of colored shapes together with an arrow cue. They were asked to describe the direction of the arrow alone (e.g., "downwards", the simple condition), the direction and shape (e.g., "downwards to the triangle", the medium condition), or the direction, shape, and color (e.g., "downwards to the grey triangle", the complex condition). Instructions on which type of phrase to produce were given before the network stimulus appeared on the screen. Marek et al. observed a higher P3 amplitude in the medium and complex conditions as compared to the simple condition. Thus, again, the ERP data suggest that speakers allocate more central capacity to the planning processes in more complex conditions.

However, an opposite effect of complexity on the P3 amplitude was observed in a study by Sikora, Roelofs, and Hermans (2016). Their participants had to produce noun phrases in response to pictures in simple (e.g., "the cup") or complex (e.g., "the green cup") conditions. Sikora et al. observed that the P3 amplitude was lower in the complex than in the simple condition, which is exactly the reverse of what Habets et al. (2008) and Marek et al. (2007) empirically obtained.

As suggested by Sikora, Roelofs, and Hermans (2016), one potential explanation of this discrepancy in results between studies was offered by Kok (2001). Based on an extensive literature review of P3 studies, Kok (2001) suggested that the P3 amplitude primarily reflects the amount of capacity allocated to "event-categorization", which refers to perceptual-conceptual processes or stimulus identification. According to Kok, attention to stimulus processing (i.e., "task emphasis") increases the P3 amplitude, while concurrent working-memory load (i.e., in dual-task performance) reduces the P3 amplitude. The amplitude is reduced because the working-memory load is thought to take away some of the capacity that is allocated to perceptual-conceptual processes. To map these factors to the present work, we refer to the capacity allocation aspect of Kok's view as "allocation of processing capacity" and the working-memory load aspect as the "use of processing capacity". Importantly, the account by Kok suggests that changes in task difficulty per se do not predict the direction of the effect on the P3 amplitude. Task difficulty will increase P3 amplitude when participants allocate more capacity on complex than on simple trials and when capacity use does not counteract this difference in allocation. In contrast, task difficulty will decrease P3 amplitude when participants allocated more capacity on complex than on simple trials but use counteracts this difference in allocation. Correspondingly, Sikora et al. explained the difference in directionality of the P3 effects between studies in terms of a difference in how much capacity was allocated and how much capacity was used when performing the task. According to their proposal, the participants of Habets et al. (2008) and Marek et al. (2007) allocated substantially more central capacity to planning processing on complex trials than on simple trials, and this difference

was preserved despite greater use of capacity on the complex trials. As a consequence, the P3 amplitude was higher on the complex trials than on the simple trials in these studies. In contrast, the participants of Sikora et al. allocated the same amount of capacity to processing on complex and simple trials but more capacity was used on complex trials, reducing the P3 amplitude. Alternatively, more capacity was allocated on complex than on simple trials, but capacity use counteracted this difference, also reducing the P3. To conclude, the difference in the directionality of the P3 effect between the studies of Habets et al. and Marek et al., on the one hand, and the study of Sikora et al., on the other, may have been due to a different balance between capacity allocation and use.

Such a different balance between capacity allocation and use may have been due to design differences between the studies. Overall, the task in the study of Sikora, Roelofs, and Hermans (2016) was simpler than the tasks in the studies of Habets et al. (2008) and Marek et al. (2007). Perhaps, the participants of Sikora et al. therefore did not bother to allocate different amounts of capacity depending on the trial type. First, the utterances produced by the participants of Sikora et al. were relatively simple compared to the utterances produced in the studies of Habets et al. and Marek et al. Second, Sikora et al. employed only a very limited stimulus set (only four pictures) presented many times throughout the experiment. In contrast, in the studies of Marek et al. and Habets et al., there were more stimuli, which were more complex and repeated less often. If more capacity is allocated on complex trials than on simple trials, this may reduce the behavioral consequences of differences in complexity. In line with this, Sikora et al. observed an effect of complexity in response time (RT), whereas Habets et al. observed no significant difference in RT and error rate between the non-chronological (complex) and chronological (simple) conditions. Marek et al. did not report mean RTs, although the complexity effect was present in the error rates.

If the difference between studies in the directionality of the P3 effect is due to an overall difference in task complexity, then we should be able to replicate the P3 findings of Habets et al. (2008) and Marek et al. (2007) by increasing task complexity using the study design of Sikora, Roelofs, and Hermans. (2016). Sikora et al. asked participants to respond to black-and-white or colored pictures by producing noun phrases with no adjectives (e.g., “the cup”, the simple condition) or with one adjective (e.g., “the green cup”, the complex condition), respectively. To increase overall task complexity in the present ERP study, we used small and large pictures and asked participants to produce noun phrases with one adjective to express size (e.g., “the small chair”, the simple condition) or with two adjectives expressing size and color (e.g., “the small red chair”, the complex condition). Furthermore, we increased the stimulus set size to a total of 16 pictures. If participants now allocate different amounts of capacity depending on the trial type, then the P3 amplitude should be larger in the complex than in the simple condition, as Habets et al. and Marek et al. observed. As in the study of

Sikora et al., the type of phrase switched every second trial. Sikora et al. did not observe an effect of switching in the amplitude of the P3, whereas a number of studies on non-linguistic performance have observed such switching effect (see Sikora et al. for a review). For example, Lorist, Klein, Nieuwenhuis, De Jong, Mulder, and Meijman (2000) observed that switching reduced the amplitude of the P3, suggesting greater use of central capacity on switch than on repeat trials. Thus, the switching manipulation in the present study may possibly index an effect of capacity use on the P3 (i.e., lowering the P3 amplitude on switch trials). Furthermore, switching has been observed to influence an earlier component, namely the N2 (e.g., Jackson, Swainson, Cunnington, & Jackson, 2001; Sikora et al.), with the ERP deflecting more negatively on switch trials than on repeat trials. Therefore, an N2 switching effect would be expected in our alternating-runs experiment.

5.2. METHODS

5.2.1. Participants

Twenty-three native Dutch speakers from the participant pool of Radboud University Nijmegen participated in this experiment for monetary compensation (16 female, M = 22.91 years, SD = 3.19 years, age range: 19-30). One participant's data were excluded from analysis because of excessive error rate: 28% of trials contained errors in naming. All participants reported having normal or corrected-to-normal vision, normal hearing, being non-dyslexic and having no speech impediments (i.e., stutter). The participants signed an informed consent form prior the experiment. The study followed the World Medical Association Declaration of Helsinki and it was performed within the line of research that was approved by the local ethics committee.

5.2.2. Materials

Sixteen line drawings of objects were selected from the picture database of the Max Planck Institute for Psycholinguistics, Nijmegen. These pictures showed high name agreement in previous studies of our lab. The depicted objects belonged to four semantic categories: human body parts, animals, kitchenware and furniture (see the Table D1 for the stimulus list). Dutch nouns have grammatical gender, namely common or neuter, and definite articles are gender-marked. The article "de" is used for nouns with common gender and "het" for nouns with neuter gender. Objects were chosen so that the gender-marked determiners associated with the object names were balanced: eight object names were associated with the determiner "de"

and the other eight with the determiner "het". Pictures were manipulated by color (white, red, blue) as well as size (small, big). Small pictures were 4-by-4 cm and large pictures were 8-by-8 cm.

In the simple condition, a picture appeared on the screen in default (white) color in either big or small size. In the complex condition, a picture appeared in either red or blue color and in either big or small size. Each picture appeared 32 times in total throughout the experiment: 16 times either big (8 times) or small (8), and 16 times either big red (4), big blue (4), small red (4), or small blue (4).

To control for trial-by-trial effects, a pseudo-randomization technique was applied using the Mix software (Van Casteren & Davis, 2006). Nouns were separated by a minimum distance of 10 trials, no initial phoneme repetition of nouns were permitted, determiners could be repeated four trials in a row maximally, and there was a minimum distance of three trials separating nouns of the same semantic category. All participants received a unique stimulus list.

5.2.3. Design and Procedure

Stimulus presentation and recording of behavioral data were implemented with Presentation Software (Neurobehavioral Systems, Albany, CA). During the experiment, participants were seated in a regularly lit sound-attenuating room, approximately 75 cm away from the screen. All vocal responses were recorded through Presentation, later to be used for error coding and response-time registration.

At the beginning of the experiment, participants were familiarized with the pictures and corresponding nouns and determiners. An experimental session consisted of eight blocks of 64 trials each. A trial began with presentation of stars centrally for 1200 ms, during which a participant was instructed to blink. This was followed by presentation of a blank (black) screen for 850 ms ± a random jitter of up to 250 ms. Then a picture was presented on the screen for 300 ms (see Figure 5.1 for the trial events). Participants were instructed to name this picture in the format "determiner + adjective(s) + noun" as quickly and accurately as possible, but not until they had planned out everything they would like to say. The picture presentation was followed by a fixation cross, presented centrally for 2500 ms, during which participants made their vocal response. Before testing, participants were given a training session consisting of ten trials with a longer time to respond (3500 ms), followed by five trials with the regular response interval (2500 ms). Following this mandatory training, participants were given the choice of practicing further (2 blocks of five trials each), after which testing began.

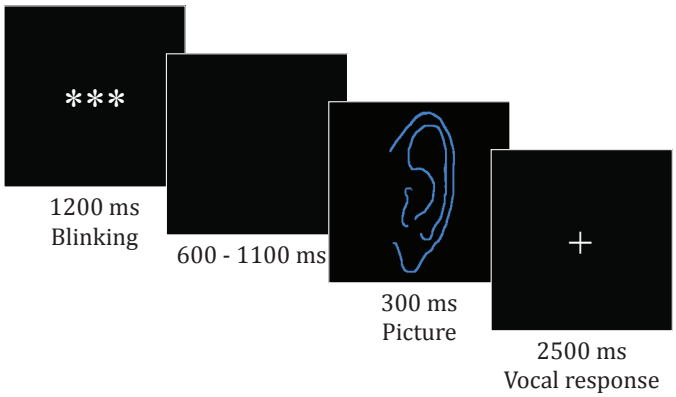

FIGURE 5.1. Trial events. After a blinking interval and a black screen, pictures of different size and color are presented for 300 ms, and participants have to respond within 2500 ms after picture offset.

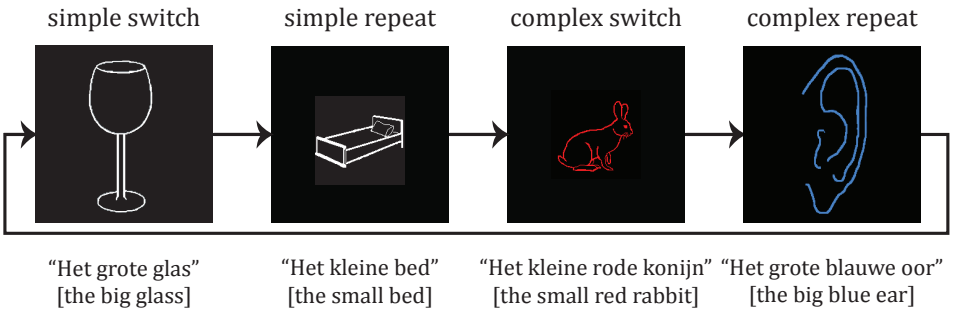

FIGURE 5.2. Illustration of the trial sequence. Following two simple trials (at which only size had to be specified in the utterance), two complex trials were presented (at which size and color had to be specified), which was followed by two simple trials, and so forth.

The trials were presented in alternating runs (Rogers & Monsell, 1995; see Figure 5.2), and the participants were informed about the sequence before the start of the experiment. Each block began with two trials of the simple condition (determiner + size adjective + noun) followed by two trials of the complex condition (determiner + size adjective + color adjective + noun). This pattern continued throughout a block of trials. On *switch* trials, the complexity level was different compared to the previous trial (i.e., a simple trial following a complex trial or a complex trial following a simple trial), whereas on *repeat* trials, the complexity level was the same between trials (i.e., a simple trial following a simple trial or a complex trial following a complex trial).

5.2.4. EEG Data Acquisition

EEG data were recorded from 32 active electrodes (ActiCAP 32Ch Standard-2, Brain Products) at a sampling rate of 1000 Hz. Data were referenced online to FCz, and then re-referenced offline to the average of the left and right mastoids. Six additional active electrodes were used in order to record eye movements and blinks (vertical and horizontal EOGs), as well as movements of the mouth (EMG). Vertical EOG was recorded from below and above the orbital rim of the left eye. Horizontal EOG was recorded from the lateral orbital rim of the left and right eye. EMG measurements were taken one above the right side of the mouth (right orbicularis oris superior), and one below the left side of the mouth (left orbicularis oris inferior). The difference of the signals recorded from these pairs of electrodes (i.e., EOG and EMG) was used for artifacts rejection. All EEG signals were online filtered with the low cutoff of 0.016 Hz and the high cutoff of 125 Hz.

5.2.5. Analysis

Trials that contained errors in naming were excluded from analysis. Errors concerned the production of an incorrect determiner, adjective or noun, or an incorrect ordering of the adjectives (size had to be expressed before color). Trials with missing determiners were not considered to be errors and were kept for analysis. After an experimental session, the recordings of the utterances were used to determine the RTs. RTs were measured semi-automatically using PRAAT (Boersma, 2002).

EEG data were analyzed using FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom analysis scripts, run through Matlab v. 8.1.0.604 (R2013a, The MathWorks, Inc.). Trials were defined to start 500 ms before the onset of a picture stimulus and finish at speech onset. To account for muscular artifacts that contaminate the EEG data close to articulation onset, trials were inspected visually and cut off at the onset of such preparatory muscular activity, usually approximately 150 ms before the speech onset. Since the average RT for the simple phrases was expected to be at least 800 ms (based on Bürki & Laganaro, 2014; Sikora, Roelofs, & Hermans, 2016), such approach would yield sufficient data to analyze the P3 component. Blinks and other eye movements were rejected manually through trial-to-trial visual inspection of vertical and horizontal EOG.

Artifact-free data were then low-pass filtered at 30 Hz and baseline-corrected per trial using a pre-stimulus baseline window of 300 ms, following previous studies that employed per-trial baseline correction in the alternating-runs paradigm (e.g., Capizzi, Ambrosini, Arbula, Mazzonetto, & Vallesi, 2016; Hung, Huang, Tsai, Chang, & Hung, 2016; Kamijo & Takeda,

2010; Sikora, Roelofs, & Hermans, 2016; Swainson, Cunnington, Jackson, Rorden, Peters, Morris, & Jackson, 2003). In order to obtain ERPs, the data were then averaged across trials per condition per participant. These averaged waveforms were analyzed using within-subjects cluster-based permutation test (Maris & Oostenveld, 2007) to assess statistical significance between conditions.

5.3. RESULTS

5.3.1. Behavioral results

Group-averaged RTs on trials manipulated by complexity and switching are displayed in Figure 5.3. Participants were slower on complex trials than on simple trials ($F(1,21) = 43.2$, $p < .01$), while there was no main effect of switching ($F(1,21) = 1.87$, $p = .19$). Complexity interacted with switching ($F(1,21) = 28.83$, $p < .01$). Naming on complex switch trials was faster than on complex repeat trials ($t(21) = 4.42$, $p < .01$), while naming on simple switch trials was slower than on simple repeat trials ($t(21) = 5.47$, $p < .01$).

Error rates for the simple repeat, simple switch, complex repeat, and complex switch trials were 2.4 %, 2.1 %, 3.3 %, and 2.5 %, respectively. The analysis of error rates revealed an only marginally significant effect of complexity ($F(1,21) = 3.37$, $p = .07$), while there was no main effect of switching ($F(1,21) = 2.27$, $p = .15$) or interaction between complexity and switching ($F(1,21) = 0.63$, $p = .44$).

5.3.2. ERP results

Group-averaged ERPs for the vertex electrode Cz are displayed in Figure 5.4 (for group-averaged ERPs over nine electrode sites, see Appendix D1). Differential time-locked EEG activity was present in three time-windows, roughly corresponding to three standard ERP components: P2 (from 150 to 250 ms post-stimulus onset), N2 (250 – 350 ms), and P3 (350 – 500 ms).

During the P2 time-window, ERP waveforms on complex trials deflected more positively than those on simple trials over a broad range of fronto-central electrode sites ($p = .002$) and more negatively over occipital electrode sites ($p = .018$). ERPs on switch and repeat trials did not differ. Complexity and switching interacted ($p = .02$), with only switch trials showing a complexity effect ($p < .001$).

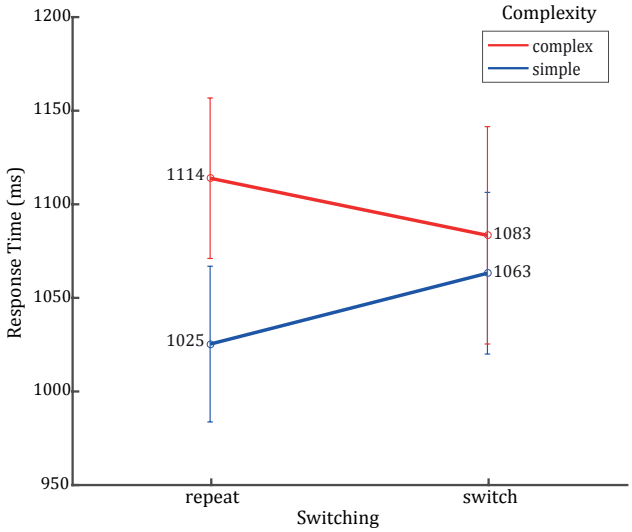

Figure 5.3. Behavioral results. Mean response time as a function of complexity (simple, complex) and switching (repeat, switch). The error bars indicate the standard error of the mean per condition.

During the N2 time-window, the main effect of complexity persisted (p = .005). Moreover, ERPs on switch trials deviated more negatively than those on repeat trials (p = .036). There was no interaction of complexity and switching.

During the P3 time-window, the complexity effect was present over centro-parietal electrode sites (p < .001), while no switching effect was obtained. However, an interaction of complexity and switching was observed (p = .037). Although complex trials differed from simple trials on both switch (p < .001) and repeat (p = .042) trials, the complexity effect on the switch trials was larger than on the repeat trials. The switching effect was only present on the simple trials (p = .049). The topographies of the simple ERP effects are shown in Figure 5.5.

As we described in the Methods section, we performed per-trial baseline-correction of the data before studying differences in ERPs. This baseline correction may be problematic when participants are certain about the upcoming stimulus type. Per-trial subtraction of averaged pre-stimulus activity indeed led to an elimination of effects of preparation in the current experiment. However, if such correction is not performed before computing ERPs, it becomes impossible to assess whether any observed effects are real or due to random fluctuations in sustained EEG activity levels across trials. Therefore, at expense of not being able to investigate preparatory effects, we chose to baseline-correct the data. Furthermore, baseline-correction is a standard practice in research employing the alternating runs paradigm, including the

study by Sikora, Roelofs, and Hermans (2016), which we used as a critical comparison for our experiment. Nevertheless, when we re-analyzed the current data without baseline-correction, the critical comparisons yielded the same results as the previously reported tests performed on the baseline-corrected data. The complex trials deflected more positively than the simple trials within the P3 time-window (p = .002) and *complexity* and *switching* interacted within the P3 time-window (p = .004), with only switch-trials showing significant complexity effect (p = .003), while no difference was observed between complex-repeat and simple-repeat trials.

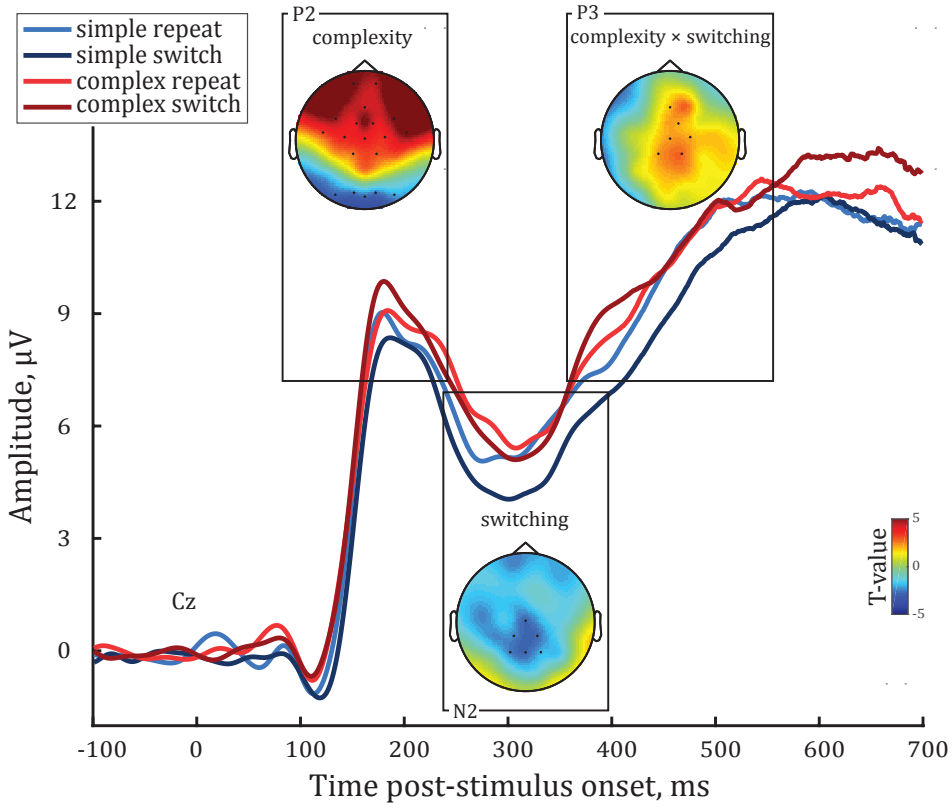

FIGURE 5.4. ERPs of simple repeat, simple switch, complex repeat, and complex switch trials. Topographies represent T-values of the cluster-based permutation tests performed on every comparison within the following time-windows: 150 – 250 ms post-stimulus onset for the P2; 250 – 350 ms post-stimulus onset for the N2; 350 – 500 ms post-stimulus onset for the P3. The complexity effect is the difference between complex and simple trials. The switching effect is the difference between switch and repeat trials. The interaction effect is a difference between complexity effects in switch and in repeat trials. Electrode sites that entered spatio-temporal clusters based on which the null hypothesis was rejected are marked.

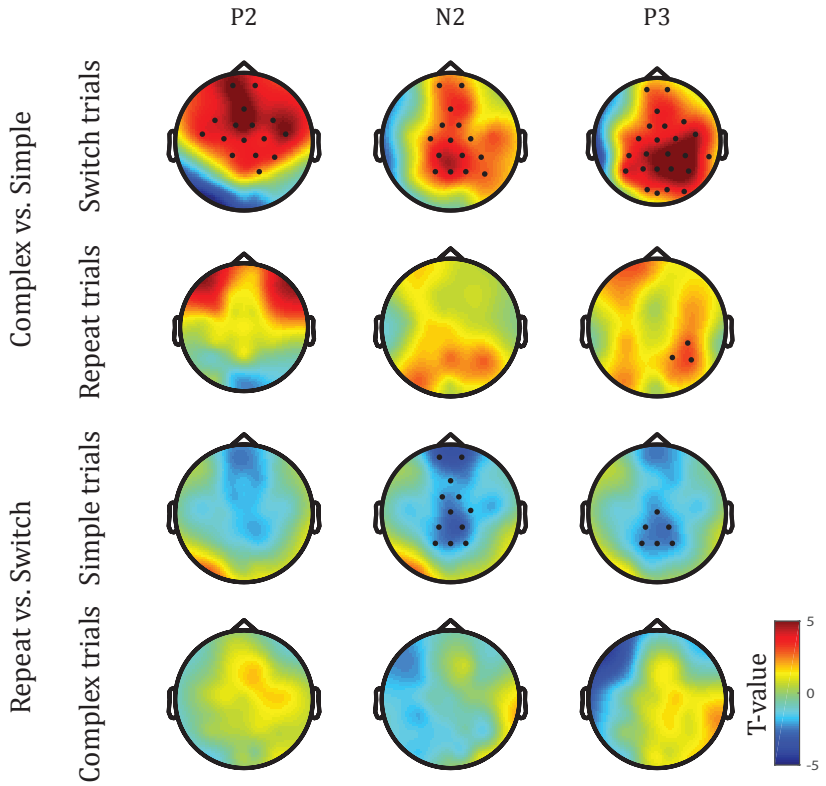

FIGURE 5.5. Topographies of simple ERP effects. Topographies represent T-values of the cluster-based permutation tests performed on every comparison. Motivated by the literature, positive-sided tests were performed on complexity effects in the P3 time-window, and negative-sided tests were performed on switching effects in the N2 and P3 time-windows. Otherwise, two-sided cluster-based permutation tests were performed on all simple effects. Electrode sites that entered spatio-temporal clusters based on which the null hypothesis was rejected are marked.

5.4. DISCUSSION

In previous ERP research on the allocation and use of processing capacity in language production, the directionality of the P3 effect differed between studies. Whereas Habets et al. (2008) and Marek et al. (2007) obtained a higher P3 amplitude for complex trials than for simple trials, Sikora, Roelofs, and Hermans (2016) observed the reverse. We hypothesized that the difference in directionality of the P3 effect may have been due to a difference in overall task complexity between studies. We tested this by increasing task complexity using the study design of Sikora et al. As in the study of Sikora et al., participants had to switch between

utterance types, which may reduce the P3 amplitude, as has been show in non-linguistic studies. The mean RTs were much longer in the present study (around 1070 ms) than in the study of Sikora et al. (around 700 ms), which suggests that our attempt to increase the overall complexity of the task was successful. In the ERPs, we observed a larger P3 amplitude on complex than on simple trials, as obtained by Habets et al. (2007) and Marek et al. (2008) but different from Sikora et al. This suggests that participants in our study allocated more capacity to the planning processes on complex than on simple trials. Also different from Sikora et al., we observed an interaction of complexity and switching, with the complexity effect being larger on switch trials than on repeat trials. An RT switch cost was present only on simple trials. This would suggest that more capacity is used in switching to simple than to complex trials, in line with the interaction between complexity and switching in the P3 amplitude.

Following Sikora, Roelofs, Hermans, and Knoors (2016), we take central processing capacity to determine "how well a speaker keeps in mind the requirements of the task (e.g., to be fast and accurate) while engaging in conceptual and linguistic processes" (p. 1720). Sikora et al. took processing capacity to concern the updating component of executive control. According to an influential theory of Miyake, Friedman, Emerson, Witzki, Howerter, and Wager (2000), three main components of executive control are updating, inhibiting, and shifting. Sikora et al. assumed that conceptual and linguistic processes critically require the updating of working memory. Long and short noun phrases differ in the extent of conceptual and linguistic processing and the corresponding demand on the updating of working memory (see Sikora et al. for discussion).

Several conceptual and linguistic processes underlie the production of noun phrases (e.g., Levelt, 1989). For example, in response to a picture of a big blue ear (Figure 5.1), participants have to conceptually identify the object as well as its size and color, and the corresponding concept representations need to be temporarily maintained in working memory. Moreover, participants have to retrieve from long-term memory the lemmas of the corresponding Dutch noun (*oor*), size adjective (*groot*), color adjective (*blauw*), and gender-marked definite article (*het*). Using these lemmas, a syntactic structure for the noun phrase has to be generated, which includes serial ordering of the determiner, size adjective, color adjective, and noun. The syntactic structure has to be maintained temporarily. Moreover, a morpho-phonological representation has to be generated and maintained, which includes retrieving the appropriate morphemes and inflecting the adjectives (e.g., *grote*; morphological encoding), retrieving the phonemes and syllabifying them (phonological encoding), and accessing articulatory programs for the syllables (phonetic encoding), followed by the initiation of articulation (for "het grote blauwe oor"). These conceptual and linguistic processes and their time course have been extensively investigated in previous RT experiments (e.g., Meyer, 1996; Schriefers, 1992) and more recently in electrophysiological experiments (e.g., Bürki & Laganaro, 2014;

Bürki, Sadat, Dubarry, & Alario, 2016; Eulitz, Hauk, & Cohen, 2000; Michel Lange, Perret, & Laganaro, 2015; Pylkkänen, Bemis, & Elorrieta, 2014). However, none of these earlier studies examined influences of capacity demands on the P3 (i.e., analyses of the P3 were not reported), which was central to the present experiment. In our experiment, conceptual and linguistic processing was presumably more extensive for the long than for the short noun phrases. As a consequence, keeping in mind the requirements of the task while engaging in conceptual and linguistic processes was more demanding for the long than for the short phrases, leading to differential allocation of central processing capacity, as reflected in the amplitude of the P3.

We obtained effects of our complexity and switching manipulations in the P2 (from 150 to 250 ms post-stimulus onset), N2 (250 – 350 ms), and P3 (350 – 500 ms). It is difficult to relate these ERP effects (or corresponding time windows) to specific conceptual and linguistic processes. In a meta-analysis of neuroimaging studies on word production, Indefrey and Levelt (2004; Indefrey, 2011) estimated that perceptual and conceptual processes in picture naming (i.e., producing a single noun) are completed around 200 ms after picture onset, while linguistic planning processes (i.e., lemma retrieval and morpho-phonological encoding up to phonetic encoding) are completed around 145 ms before articulation onset. These estimates hold for a mean picture naming RT of 600 ms. However, in the present experiment, the conceptual and linguistic processes concerned the planning of a phrase containing three or four words (i.e., a determiner, one or two adjectives, and a noun) rather than a single noun. The mean RT in the present experiment was around 1070 ms. Clearly, the estimates of Indefrey and Levelt for single word production cannot be directly applied to noun phrase production. At the very least, the estimates have to be rescaled (Roelofs & Shitova, 2017). Under proportional rescaling (i.e., 1070/600 × 200 ms), the estimate for the completion of perceptual and conceptual processes in the present experiment would be about 360 ms. This would imply that the observed effects in the P2 and N2 occurred during perceptual and conceptual encoding, whereas the effects in the P3 arose during linguistic planning processes.

In previous ERP studies of picture naming, Strijkers, Costa, and Thierry (2010) and Strijkers, Holcomb, and Costa (2011) argued that the P2 reflects the "intention to speak". The time windows of the P2 in these studies were 160-240 ms and 140-210 ms, and the mean picture naming RTs were about 720 ms and 800 ms, respectively. Under proportional scaling of the estimates of Indefrey and Levelt (2004; Indefrey, 2011), completion of perceptual and conceptual processing was, respectively, around 240 ms and 266 ms in these studies. This would suggest that the P2 effects in the studies of Strijkers et al. occurred during perceptual and conceptual processing.

In our study, complexity effects were present in the P2, N2, and P3 components of the ERP. This is not surprising, since complex and simple trials differed at all stages of stimulus processing from visual perception (colored vs. non-colored stimuli) to response planning (expressing size and color vs. size only). A switching effect was only present in the N2 component. Moreover, complexity and switching interacted in the P2 and P3 components.

The P2 component has been associated with late visual processing modulated by salience and attention (see Luck & Hillyard, 1994, for a color pop-out P2 effect). Thus, a P2 complexity effect in our paradigm is expected: Colored stimuli were more salient events in a stream of colored and non-colored stimuli. Moreover, the color of the colored stimuli was relevant for the response, and therefore the colored stimuli triggered a larger P2 component. Interestingly, this complexity effect was present on switch trials but not on repeat trials. This can be explained by the fact that the visual pattern of the stimuli was changed on switch trials but remained the same on repeat trials. There were no early visual P1 effects associated with our color manipulation. Martinovic, Gruber, and Müller (2008) showed that such effects occur in comparisons of objects pictured in prototypical or non-prototypical colors, as well as if objects differ in amount of surface detail and linear complexity. The line drawings that we employed in the current study for complex and simple conditions were not manipulated in these dimensions and, therefore, did not differ in their ease of recognition, which explains absence of the P1 effect. Although no canonical N1 effect was observed in the current study, there was a negative-going complexity effect over the occipital-most electrode sites that temporally overlapped with the frontal-central-parietal P2 complexity effect. In particular, the P2 on Cz corresponds to an N170 on posterior sites. We have marked out the relevant time window on posterior channels in the figure D1. Therefore, the observed P2-time-window effect can be more broadly considered as an N1-P2 complex effect. This would not change the original interpretation of the P2 effect in terms of perceptual processing regulated by attention (see Proverbio, Burco, del Zotto, & Zani, 2004, for an N1 effect in color manipulation associated with attention).

The N2 component has been associated with inhibition and strategic monitoring (Folstein & Van Petten, 2008). In task switching, the amplitude of the N2 tends to be larger on switch than repeat trials (e.g., Jackson, Swainson, Cunnington, & Jackson, 2001). In our study, the ERPs were also more negative-going on switch trials than on repeat trials, replicating the pattern reported by Sikora, Roelofs, and Hermans (2016). The authors also observed an interaction of complexity and switching in the N2 amplitude, with a switch effect only being present on simple trials. Sikora et al. argued that colored pictures require long-phrase responses (e.g., "the red cup") but afford a short-phrase response as well (e.g., "the cup"), and therefore the task set for short phrases needs to be inhibited. As a consequence, later switching to a short-phrase response requires overcoming the previous inhibition, which increases the N2. There

was no interaction of complexity and switching in the N2 time-window in the current data set, but an additional post-hoc analysis showed that a switch effect was present on simple trials ($p = .017$), but not on complex trials. According to Sikora et al., overcoming the inhibition on short switch trials explains the switch cost in RTs on the short trials. However, in our RTs, both a switch cost on the simple trials and a switch benefit on the complex trials were present. One possible explanation of the switch benefit is that this effect might be due to additional lexical competition with a different color adjective produced on the previous trial, which may have occurred on a proportion of the complex repeat trials (e.g., "the big blue ear" following "the small red desk"). However, a post-hoc analysis of complex repeat trials showed no RT difference between trials that repeated the color adjective and trials that did not. Therefore, the switch benefit effect in RTs remains to be explored further in future research.

It may also be argued that the asymmetrical switch cost is due to a potential need to suppress the adjective *white* when responding to the simple stimulus and particularly after using a color adjective on the previous complex trial. However, we believe that this was not the case. First, if such suppression had to be performed on every simple-switch trial, then there should have been considerable numbers of errors that include the word *white*, which we did not observe. Second, the suppression should have led to increased RTs for simple trials compared to complex trials, on which the participant did not have to suppress the color name. However, we observed that mean RTs were about 55 ms smaller on simple than on complex trials.

Complexity and switching interacted in the P3 time-window: The P3 effect of complexity was larger on switch trials than on repeat trials. The RTs suggest that processing on the short switch and the complex repeat trials was more demanding than processing on their switching counterparts (i.e., the short repeat and the complex switch trials). This relative increase in RTs parallels the relative decrease in P3 amplitude on these trials, suggesting that additional use of processing capacity decreased the P3 amplitude. Therefore, although in general complex trials were associated with a higher P3 amplitude than simple trials, reflecting a difference in allocation of processing capacity, the greater use of capacity on short switch and complex repeat trials reduced the P3. This differential allocation and use of processing capacity explains the interaction of complexity and switching that we observed in the P3 time-window.

To conclude, we observed that allocation and use of processing capacity differentially influence the P3 amplitude in phrase production. Thus, our study demonstrated both directionalities of the P3 effect proposed by Kok (2001).

CHAPTER 6

GENERAL DISCUSSION

6.1. SUMMARY AND EVALUATION

This dissertation explored various aspects of competition and adjustment in language production using electrophysiological and behavioral methods. First, I provided new evidence for competitive lexical selection in spoken word production (in the Stroop and PWI tasks). Then I explored circumstances in which control over word planning can be strategically adjusted depending on participants' expectations regarding upcoming task. Finally, I presented more general evidence for adjustments in processing capacity in a phrase-production task.

The study reported in Chapter 2 directly addressed the question of the locus of interference effect in the PWI task as compared to the Stroop task. The competitive account (Levelt et al., 1999; Roelofs, 2014) explains both PWI and Stroop effects through varying degrees of competition that a target word needs to overcome in order to be selected. An alternative account (e.g., Dell'Acqua et al., 2010; Van Maanen et al., 2009) holds that the competition account is only valid for the Stroop task, while the PWI effect arises from a different mechanism. Because of different nature of the stimuli (pictures in PWI vs. colors in Stroop), instead of occurring at the stage of lexical selection, the PWI effect is associated with the early stage of perceptual encoding and conceptualization. In the study presented in Chapter 2, the Stroop-like effects in the Stroop and PWI tasks were compared directly using the same group of participants and experimental parameters (e.g., number of trials, trial definition). The Stroop-like effects in both Stroop and PWI were associated with the N400 ERP effect, while no earlier effects were present, which would link to perceptual encoding and conceptualization. Importantly, there was no systematic difference in timing of the N400 onset in the Stroop task and a variant of the PWI task, where the number of stimuli was controlled for. These results suggested that the functional locus of the Stroop-like effect is linked to the lexical selection stage of the processing stream, independent of the stimulus type. This pattern of results provides new evidence for the competitive account.

In Chapter 3 I further elaborated on the functional locus of the PWI effect and explored it from the perspective of control adjustments in a word production task. The effects of anticipation of a certain stimulus type were studied extensively using variations of the Stroop task, but only a few studies so far researched this topic from the point of view of picture naming. I presented a PWI study where the Stroop-like effect was assessed as a function of the previous trial type: It was expected that, due to control adjustments, if the previous trial was incongruent the Stroop-like effect would be smaller than if the previous trial was congruent (i.e., the Gratton effect). In line with the previous study, reported in Chapter 2, the Stroop-like effect was observed during the word planning stage (based on the timing of the N400 effect). Importantly, this was true for both post-congruent and post-incongruent trials, with no systematic difference in timing of the N400 onset. These results are in contrast with one

previous study (Van Maanen & Van Rijn, 2010) that reported shifts in the functional locus of the Stroop-like effect in PWI depending on the previous trial type. The research presented in Chapter 3 confirmed that the competitive account holds independently of experienced conflict on the previous trial. The Gratton effect was observed in RTs, N400 amplitude, and mid-frontal theta power, suggesting that there are indeed differences in stimulus processing on post-congruent and post-incongruent trials. However, these differences were quantitative and there was no evidence for qualitative shift in the locus of the Stroop-like effect.

The control adjustments in picture naming were further explored in Chapter 4, where I focused on the nature of the Gratton effect in speaking. There exist two explanations of this phenomenon. A dominant account (Botvinick et al., 2001) holds that decrease of the Stroop-like effect on post-incongruent trials, as compared to post-congruent trials, is due to the experienced conflict on incongruent trials. An alternative account (Lamers & Roelofs, 2011) suggests that such behavioral pattern is rather due to expectation of congruency. It predicts that the Stroop-like effect is larger on post-congruent trials than on post-incongruent and, importantly, post-neutral trials. In a series of three behavioral experiments control adjustments were studied in single-word as well as phrase-production versions of the PWI task. The Stroop-like effect on post-congruent was significantly larger than on post-neutral trials, while the effects on post-neutral and post-incongruent trials did not differ. These results provide support to the alternative account that maintains that the control adjustments are due to expected congruency rather than post-conflict adaptation of processing.

Chapter 5 takes the topic of adjustments in phrase production to a more general level. Here, adjustment effects on complexity and switching were assessed. The presented study demonstrated that speakers can adjust allocation of processing capacity based on available information about the complexity of an upcoming trial. Moreover, such adjustment of amount of allocated processing capacity is reflected in the same P3 ERP component that shows consumption of processing capacity.

To conclude, in a series of studies I examined the competitive account in single-word and phrase production tasks. More specifically, I presented evidence in favor of the model of language production that maintains that the distractor effect in the Stroop-like tasks, including PWI, is linked to the word-planning stage of the processing stream. On the contrary, I found no evidence for an earlier functional locus of the PWI effect, as suggested by alternative accounts. Furthermore, I investigated how subjects' expectations regarding upcoming trial type influence their strategies on the task. I presented research on adjustments of control over word planning in a PWI task with trial-by-trial randomization and on adjustments of capacity allocation in a phrase-production picture naming task using alternating runs. Research on the nature of control adjustments in PWI, presented in this dissertation, demonstrated that

such adjustments are a result of expectation of congruency and primarily driven by post-congruent trials, not by post-incongruent trials, as suggested by the currently-dominant conflict-monitoring account. Moreover, analysis of brain activity patterns associated with control adjustment in PWI suggested that the locus of the Stroop-like effect is fixed, linked to the word-planning stage, providing new evidence to the competitive account. Finally, in the last study I discussed another type of adjustments – capacity allocation. Different from the previously described situation with the PWI task, the participants were given different stimulus types and could adjust their capacity allocation to optimally process a more complex or a less complex picture, as well as prepare a more elaborated or less elaborated response. The results showed that the ability to adjust to the stimulus complexity influenced subjects' performance both at the stage of perceptual encoding and conceptualization (i.e., the early stage of picture processing) and at the stage of linguistic processing.

In this dissertation I only present a few cases of expectancy-based control adjustments in language production. It would be interesting to examine whether control adjustment effects in terms of the N400 amplitude and/or midfrontal theta power can also be obtained in other manipulations (e.g., in experiments with probabilistic cues or with variable global probability of stimuli of different types). Given the similarity of such adjustments to the control adjustments pattern discussed in Chapters 3 and 4 (Aarts & Roelofs, 2011; Gratton et al., 1992), it appears quite likely that our results will be replicated with such designs.

More electrophysiological data on control adjustments in overt speech would allow to better understand if the Gratton-like patterns that we observed in the N400 amplitude and midfrontal theta power are underpinned by the same brain dynamics or not. In Chapter 3 (section 3.4.2) we discuss the available evidence for this question, which provisionally makes a case that the N400 and the midfrontal theta power effects result from the same brain process. On the other hand, according to the only systematic comparison of the N400 and midfrontal oscillatory activity in PWI (Piai, Roelofs, Jensen, et al., 2014), these two patterns of brain activity can be functionally distinct, being associated with different brain sources and condition orderings in comparisons including congruent, incongruent, and unrelated trials. Piai et al. suggested that the midfrontal theta-alpha power modulations reflect "different degrees of effort in resolving the competition among the alternative words, as reflected in the RTs", while the evoked activity (i.e., the N400) reflects "different degrees of effort (priming) in processing the candidate words activated by the picture-word stimuli". With respect to this functional differentiation between the N400 and the midfrontal theta power, it appears of special interest to investigate the nature of the Gratton effect using MEG, in order to compare patterns of N400 and theta power results in post-congruent, post-incongruent, and post-neutral trials. If the interpretation suggested by Piai et al. regarding different functional roles of midfrontal theta and N400 is correct, the theta power pattern should reflect the RT

results which I presented in Chapter 4 of this dissertation, since both theta power and RT reflect conflict resolution. Therefore, there will be a larger theta power Stroop-like effect on post-congruent trials than on post-incongruent and post-neutral trials, while no difference in the effect size between post-incongruent and post-neutral trials is expected.

The fact that electrophysiological patterns of brain activity are expected (or not expected) to match the RT patterns deserves extra clarification. Electrophysiology registers ongoing brain activity; RTs reflect the final outcome of multiple processes that occur during stimulus processing. Therefore, an absence of such match should not be taken to devaluate the data. Moreover, while interpretation of RT effects is usually straightforward (i.e., a larger RT value is associated with a more difficult task), we cannot intuitively evaluate electrophysiological data and we can only view the output of serious data reduction (usually, averages over multiple trials/channels/subjects) in light of other similar research. Furthermore, the output of comparison of M/EEG effects and RTs can be influenced by parameters of M/EEG data processing, use of baselining, ways of operationalizing the effects in question (see Freund et al., 2016; for discussion on why taking a simple difference in RT values (in their case) might be a suboptimal way to describe an effect). Lastly, we need to define what output of comparison of electrophysiological pattern and RT pattern we consider a match and a mismatch. For example, in the study described in Chapter 3, Gratton-like decreases of the Stroop-like effect on post-incongruent relative to post-congruent trials were observed in RTs and N400 amplitude. This was taken as a confirmation that the N400 and RTs reflect the same process of various degrees of competition in the PWI task. However, as we discussed (see 3.4.2), the Gratton effect in RTs was mostly due to longer responses on congruent-after-incongruent trials, as compared to congruent-after-congruent trials, while current-incongruent trials did not 'contribute' much to the observed interaction. On the contrary, the Gratton-like interaction in the N400 amplitude was mostly supported by decreased values on incongruent-after-incongruent trials, compared to incongruent-after-congruent trials, while the N400 amplitude on congruent trials did not appear to depend on the previous trials. Therefore, our interpretation of correspondence between RT patterns and ongoing brain activity as measured by M/EEG depends on the way we define matching patterns. To my knowledge, there exist only a few studies that directly assessed correlation between RTs and ongoing brain activity (Mars et al., 2008; Cohen & Cavanagh, 2011), therefore, methodological basis of such investigations remains unclear (e.g., whether and how electrophysiological data should be normalized for such comparisons).

6.2. ON METHODS OF MUSCLE ARTIFACTS REJECTION IN LANGUAGE PRODUCTION RESEARCH

Another methodological problem that should be considered with respect to this dissertation is use of electrophysiological techniques in language production studies and, more specifically, dealing with articulation artifacts that naturally contaminate the M/EEG recordings. It has been known for many years now that movements of lips, tongue, and jaw, as well as corresponding activity of the facial muscles can produce large-scale distortions in the EEG signal (e.g., Brooker & Donald, 1980). However, the field of language production still did not adopt a conventional way of addressing the problem of speech-related artifacts in neurocognitive experiments in language production. Two approaches that are employed relatively often in the literature are low-pass filtering and various applications of blind source separation (BSS). The first one operates under the assumption that speech-related muscular artifacts are only present in the top part of the frequency spectrum of the EEG signal (usually taken as above 30 Hz), while the lower part of the frequency spectrum remains clear from artifacts. This assumption was proven wrong in a number of studies: The muscular artifacts appeared to be present in the lower portion of the frequency spectrum (e.g., Goncharova, McFarland, Vaughan, & Wolpaw, 2003). Moreover, heavy low-pass filtering of data can distort early 'cognitive' components of the signal, while not accounting for slow drift-like artifacts, for instance, produced by jaw movements (e.g., Porcaro, Medaglia, & Krott, 2015; for how low-pass filtering compares to other methods of speech artifacts rejection). Another standard approach relies on BSS techniques followed by rejection of some spatio-temporal components selected based on certain criteria. These criteria can rely solely on quantitative characteristic of a component (e.g., kurtosis, entropy, autocorrelation, frequency composition) or they could take into consideration whether the time course of a component is correlated with a direct recording from above a muscle that potentially contributes to the speech artifact (e.g., De Vos et al., 2010, Porcaro et al., 2015).

I see three major problems with applications of the BSS approach. First, use of a BSS technique in order to identify cognitive-EEG vs. muscle-EMG components in a mixed scalp EEG signal (e.g., canonical correlation analysis, De Clercq, Vergult, Vanrumste, Van Paesschen, & Van Huffel, 2006; De Vos et al., 2010; independent component analysis, Porcaro et al., 2015) presupposes that the cognitive-EEG portion in no way determines appearance of muscle-EMG portion. This assumption certainly does not hold in language production experiments, where the brain activity from the stimulus onset (or even earlier) is targeted at planning, preparing, and producing a response (This caveat was discussed in personal communication with Vitória Piai). This shortcoming seems to be largely due to the fact that the BSS-based techniques of speech artifacts removal were often inspired by similar mechanisms that are conventionally used for eye blinks and movements or electromyogenic bursts, which indeed

can appear independently of the task. In connection to this, current BSS-based algorithms of speech artifacts removal do not appear to use information of the speech onset, which is usually readily available in language production studies as behavioral part of experimental paradigm. Lastly, application of criteria of retention/rejection of components remains largely arbitrary, whether it is done automatically using thresholds (customized to the data) or through visual examination of topographies or time courses of component data. Furthermore, evaluation of methods of artifacts rejection can only rely on how well artifactual activity was suppressed (see Ouyang et al., 2016; for discussion), since the amount of superfluously rejected brain activity can rarely be estimated (see Barbati, Porcaro, Zappasodi, Rossini, & Tecchio, 2004; for a method of recovering neural data after excessive artifacts rejection; see Cohen, 2014, Chapter 8; for discussion on strategies of rejection of independent components). A new method of speech artifacts rejection was recently proposed by Ouyang, Sommer, and Zhou (2015a, 2015b; see also Ouyang et al., 2016), which appears to address some of the shortcomings of BSS-based techniques and to outperform them in a direct comparison (Ouyang et al., 2016). The method is based on residue iteration decomposition (RIDE), that estimates artifactual activity as a component of the EEG scalp signal that is strongly locked to the response onset. It still remains to be seen whether this method allows to separate late neuronal components, e.g. associated with articulatory programming, from muscular artifacts without spuriously rejecting response-locked neuronal activity.

In all three EEG studies presented in this dissertation an alternative approach to speech artifacts was adopted. First, we measured exact RTs per trial, based on visual inspection of oscillogram and time-frequency representation of per-trial recordings of participants' responses (refer to Kessler, Treiman, & Mullennix, 2002; for common problems with voicekey measurements). EEG trials were cut off the continuous data relative to the stimulus onset (500 ms before it) and to the speech onset (right at speech onset in Chapters 3 and 5, 100 ms before speech onset in Chapter 2). Then EEG data from all electrode sites including brain data as well as peri-oral EMG time series were loaded per trial for visual inspection. Articulation artifacts were identified as high-frequency activity that occurred simultaneously close to the end of the trial over peri-oral EMG as well as a range of brain EEG electrodes. The portions of trial data which were identified as artifactual were removed from the data. Articulation artifacts could also appear in other forms (e.g., slow drifts) or in other time-windows (e.g., in the beginning of a trial). In the first case such trials were removed independently of the time-window when it occurred. In the second case the trial was marked as artifact and removed from data analysis completely. Although resourceful, this method allowed to achieve three main goals that we pursued. First, it allowed to keep as much clean data as possible for the analysis (compared to a method where a common cut-off point is determined for all trials). This was particularly important for studies reported in Chapters 3 and 5, since we assessed rather late effects there and performed time-frequency analysis of slow theta-band oscillations. Second, the method

we adopted included per-trial examination of data, which meant efficient ways of dealing with artifacts of unclear nature, such as EEG drifts. Finally, this method did not involve any type of data manipulation (which is the case for BSS-based methods of articulation artifacts rejection), i.e. we could be certain that none of the true characteristics were missing and none of the spurious effects were present in the EEG data corrected for speech artifacts.

One potential downside of the method we suggested is unequal contribution of shorter and longer trials. The tails of trials are usually contaminated with speech artifacts and excluded from analysis, therefore, the beginnings and the ends of the averaged ERPs or time-frequency representations are computed based on different numbers of trials (Figures A1 and B1). Moreover, the tails of these averages are dominated by longer trials which had longer RTs. Furthermore, averaged ERPs or TFRs of experimental conditions associated with longer RTs (e.g., incongruent trials in a Stroop-like task) consist of greater number of sample points than those of experimental conditions associated with shorter RTs (e.g., congruent trials in a Stroop-like task). This discrepancy is intrinsic to the method, but it raises a bigger question that the language production research still needs to answer: When RTs vary so much within and between experimental conditions and within and between participants, how valid is it to perform condition-wise and participant-wise averaging of the EEG signal? And how valid is it to further compare the averages in the same time scale without performing any correction of latency of components on the EEG signal? Possible solutions might include dynamic time warping transformation (Casarotto, Bianchi, Cerutti, & Chiarenza, 2005; Huang & Jansen, 1985), the RIDE method (Ouyang et al., 2016), already mentioned here in the discussion of articulation artifacts rejection, or spacio-temporal (topographic) analysis of ERPs in terms of microstates of various durations (e.g., Bürki & Laganaro, 2014).

6.3. CONCLUSIONS

In the four studies included in this dissertation I investigated whether in language production words are selected by competition among alternatives, as well as whether the process of response planning can be affected by adjustments in control or capacity allocation. In spite of a large body of available evidence in support of the competitive account, alternative explanations for established cognitive effects are still advanced that challenge the competition hypothesis. In this dissertation I addressed the question whether the competitive account can equally well explain the Stroop-like effect in language production tasks with different stimulus materials (i.e., pictures and colors) and with different experimental design parameters (i.e., number of stimuli and number of stimulus repetitions). Independently of stimulus type and design parameters, the competitive hypothesis was confirmed, while no evidence for the alternative pre-lexical locus was obtained. In the following studies included

in this dissertation I investigated various types of adjustments that can influence word or phrase planning. I argued that manipulations of participant's expectancy of the upcoming stimulus type influences their performance, as well as the neuronal patterns associated with stimulus processing and response planning. Investigation of control adjustments in word and phrase production showed that the competition hypothesis holds independently of the amount of control a participant exerts on a trial, while adjustment-driven changes in the stimulus processing stream only concern degree of effort or efficiency of processing, but not the functional locus of the interference effect in a Stroop-like task.

APPENDIX A. SUPPLEMENTARY MATERIALS FOR CHAPTER 2

TABLE A1. Stimulus materials

Task	Category	Color/picture name	Incongruent distractor
Stroop	Colors	rood (red, 49)	blauw
		blauw (blue, 33)	groen
		groen (green, 28)	rood
Stroop-like PWI	Animals	konijn (rabbit, 19)	varken
		varken (pig, 25)	beer
		beer (bear, 25)	konijn
Standard PWI	Kitchenware	pan (pan, 9)	bord
		bord (plate, 27)	fles
		fles (bottle, 48)	pan
	Clothing	trui (sweater, 12)	broek
		broek (trousers, 67)	jas
		jas (jacket, 48)	trui
	Transportation	bus (bus, 65)	trein
		trein (train, 73)	vliegtuig
		vliegtuig (airplane, 90)	bus
	Buildings	toren (tower, 19)	fabriek
		fabriek (factory, 21)	kasteel
		kasteel (castle, 28)	toren
	Body parts	neus (nose, 70)	voet
		voet (foot, 51)	oog
		oog (eye, 68)	neus
	Fruit	peer (pear, 2)	banaan
		banaan (banana, 5)	appel
		appel (apple, 10)	peer
	Food	worst (sausage, 9)	kaas
		kaas (cheese, 23)	brood
		brood (bread, 34)	worst
	Birds	kalkoen (turkey, 10)	duif
		duif (pigeon, 5)	haan
		haan (roster, 4)	kalkoen
	Cutlery	lepel (spoon, 5)	vork
		vork (fork, 5)	mes

TABLE A1. Continued

Task	Category	Color/picture name	Incongruent distractor
		mes (knife, 46)	lepel
	Tools	hamer (hammer, 9)	tang
		tang (pliers, 4)	zaag
		zaag (saw, 4)	hamer
	Building parts	dak (roof, 55)	trap
		trap (stairs, 52)	muur
		muur (wall, 67)	dak
	Bags	rugzak (backpack, 8)	koffer
		koffer (suitcase, 34)	tas
		tas (bag, 58)	rugzak
	Furniture	kast (cupboard, 30)	stoel
		stoel (chair, 51)	tafel
		tafel (table, 83)	kast

English translations and word frequency per million words in parentheses.

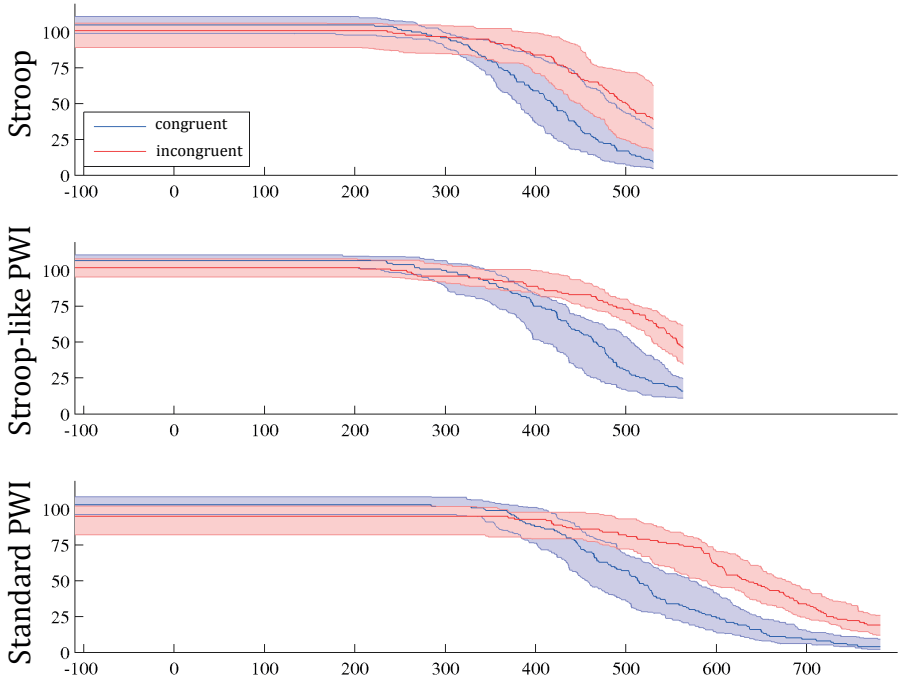

FIGURE A1. Median number of trials averaged for individual ERPs

The shaded area marks the 25th and 75th percentiles.

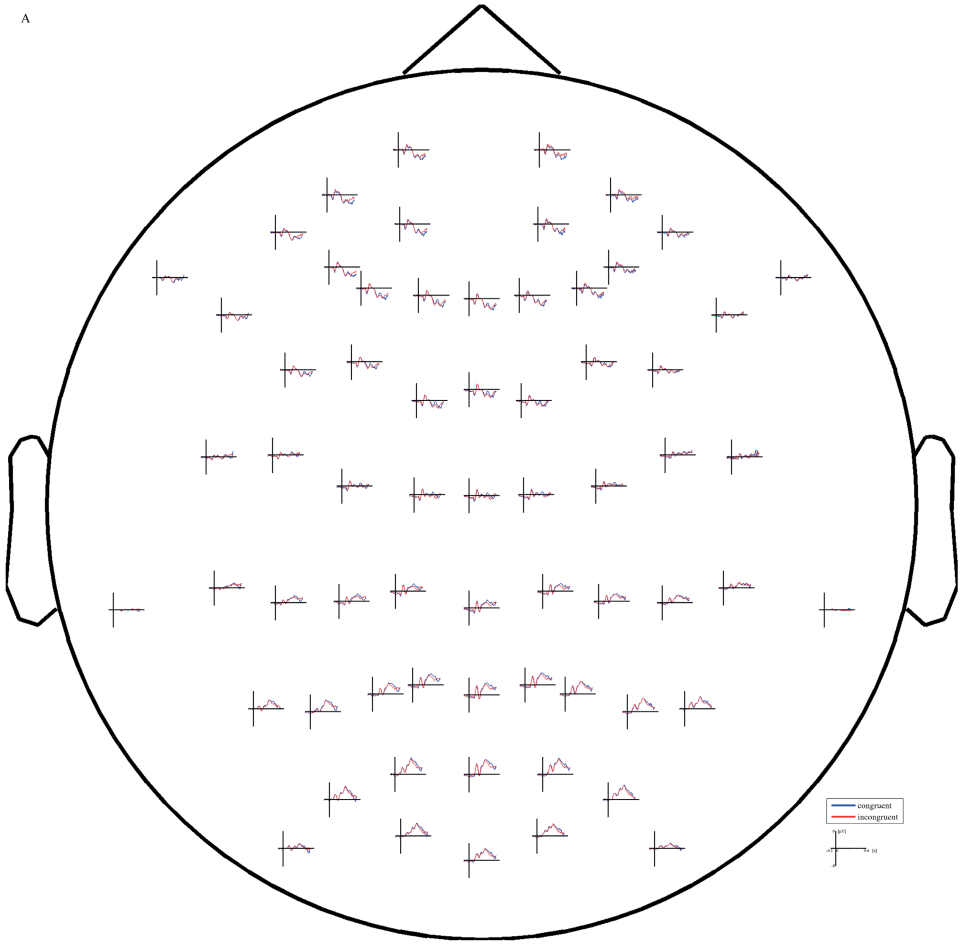

FIGURE A2. Group-averaged ERPs

(A) The Stroop task. (B) The Stroop-like PWI task. (C) The Standard PWI task.

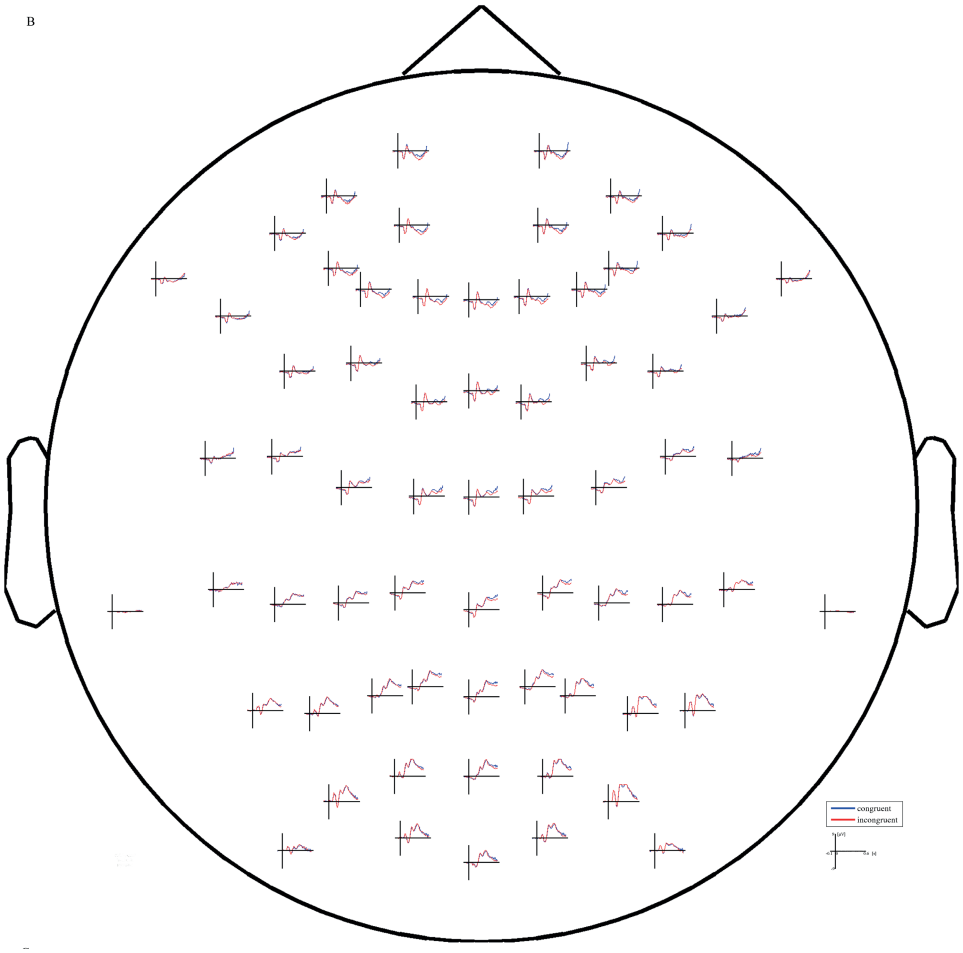

FIGURE A2. Group-averaged ERPs

(A) The Stroop task. (B) The Stroop-like PWI task. (C) The Standard PWI task.

C

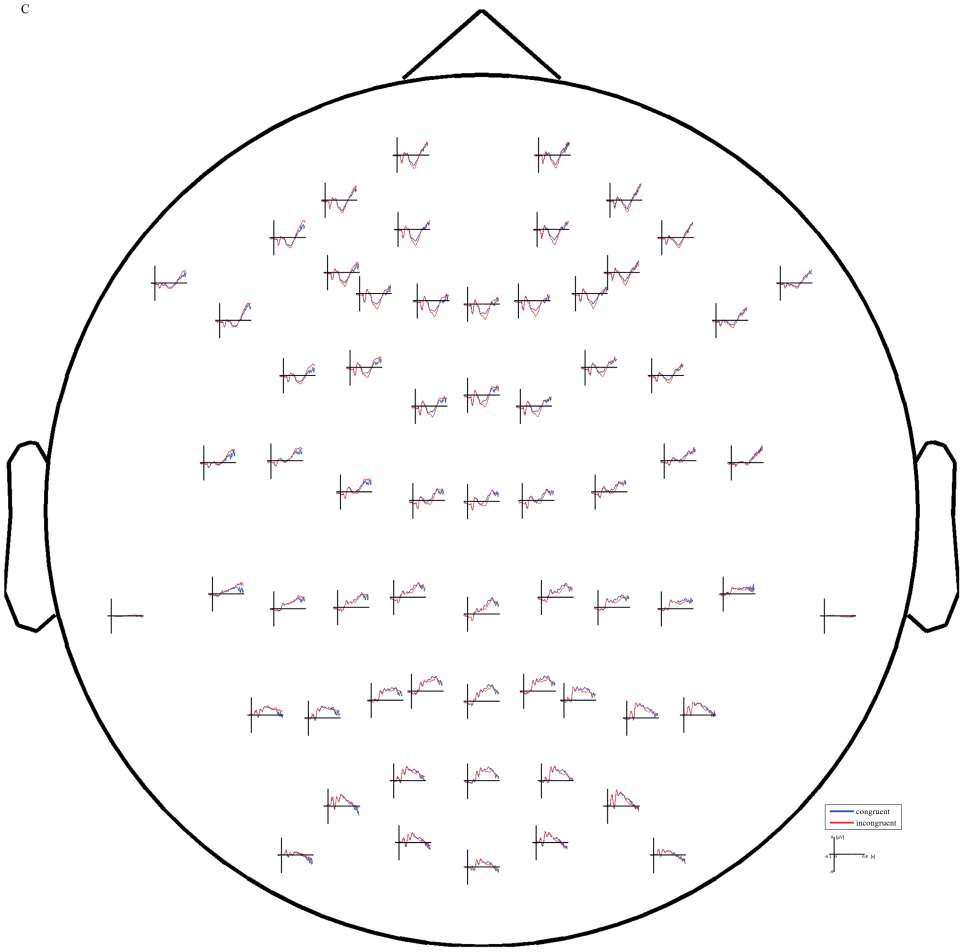

FIGURE A2. Group-averaged ERPs

(A) The Stroop task. (B) The Stroop-like PWI task. (C) The Standard PWI task.

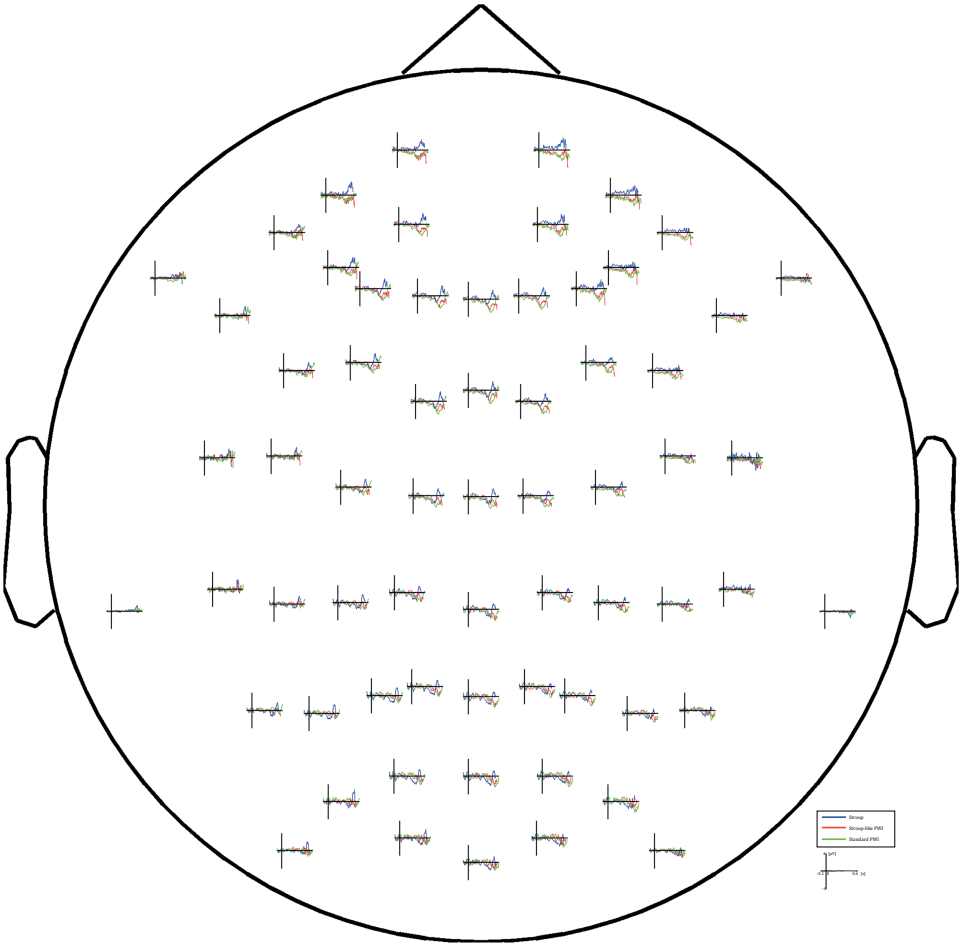

FIGURE A3. Difference waveforms for the three tasks

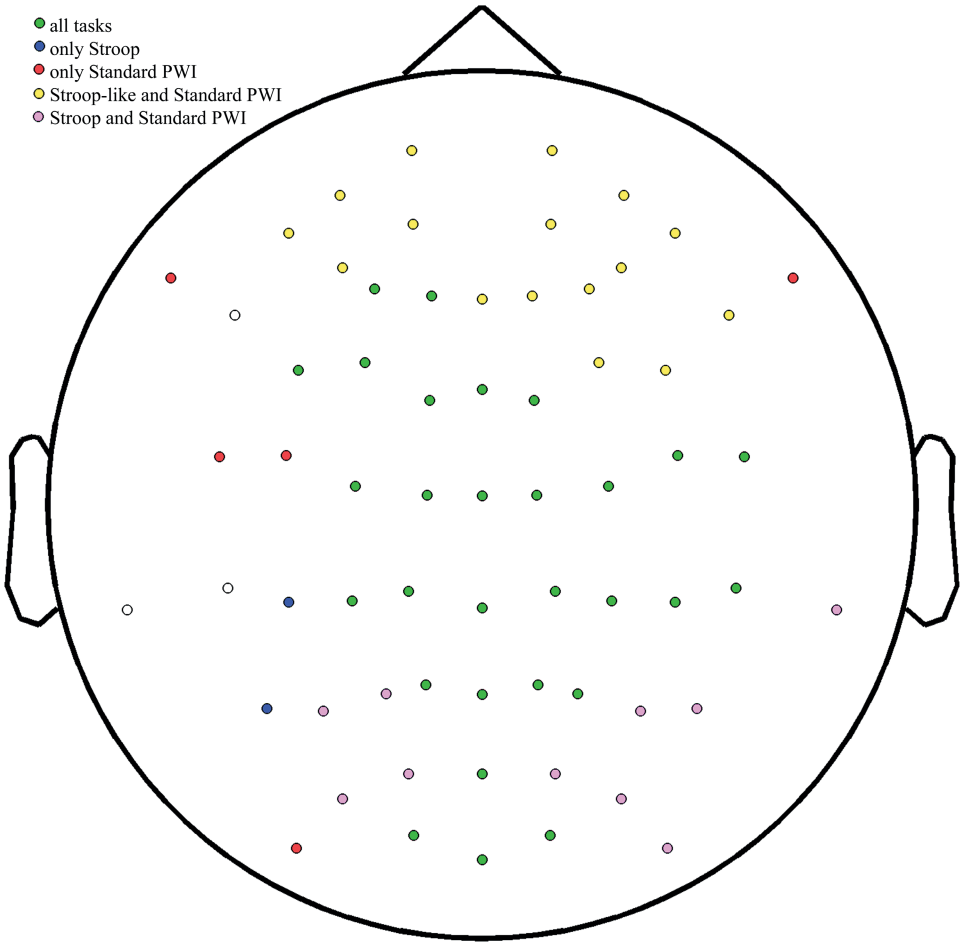

FIGURE A4. Electrode sites of the spatio-temporal clusters that showed significant differences between congruent and incongruent trials

APPENDIX B. SUPPLEMENTARY MATERIALS FOR CHAPTER 3

TABLE B1. Stimulus materials

Category	Picture name	Incongruent distractor
Kitchenware	pan (pan)	bord
	bord (plate)	fles
	fles (bottle)	pan
Clothing	trui (sweater)	broek
	broek (trousers)	jas
	jas (jacket)	trui
Transportation	bus (bus)	trein
	trein (train)	vliegtuig
	vliegtuig (airplane)	bus
Buildings	toren (tower)	fabriek
	fabriek (factory)	kasteel
	kasteel (castle)	toren
Body parts	neus (nose)	voet
	voet (foot)	oog
	oog (eye)	neus
Fruit	peer (pear)	banaan
	banaan (banana)	appel
	appel (apple)	peer
Food	worst (sausage)	kaas
	kaas (cheese)	brood
	brood (bread)	worst
Birds	kalkoen (turkey)	duif
	duif (pigeon)	haan
	haan (roster)	kalkoen
Cutlery	lepel (spoon)	vork
	vork (fork)	mes
	mes (knife)	lepel
Tools	hamer (hammer)	tang
	tang (pliers)	zaag
	zaag (saw)	hamer

TABLE B1. Continued

Category	Picture name	Incongruent distractor
Building parts	dak (roof)	trap
	trap (stairs)	muur
	muur (wall)	dak
Bags	rugzak (backpack)	koffer
	koffer (suitcase)	tas
	tas (bag)	rugzak
Furniture	kast (cupboard)	stoel
	stoel (chair)	tafel
	tafel (table)	kast

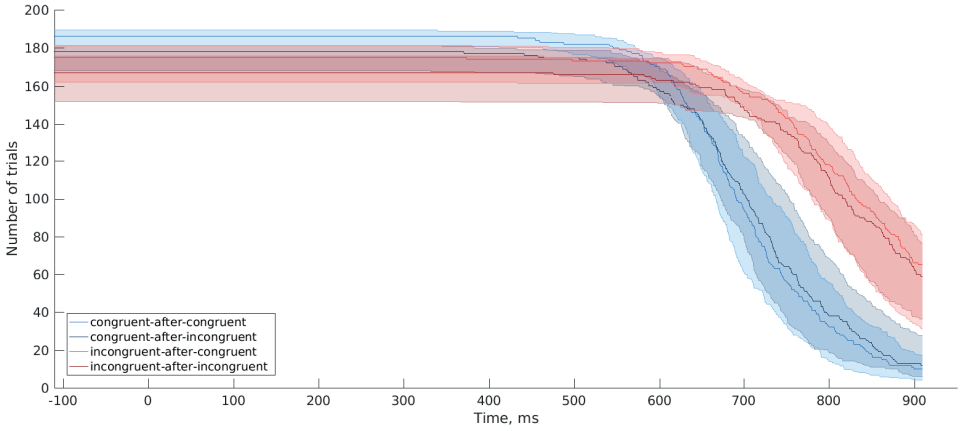

FIGURE B1. Median number of trials averaged for individual ERPs

The shaded area marks the 25th and 75th percentiles.

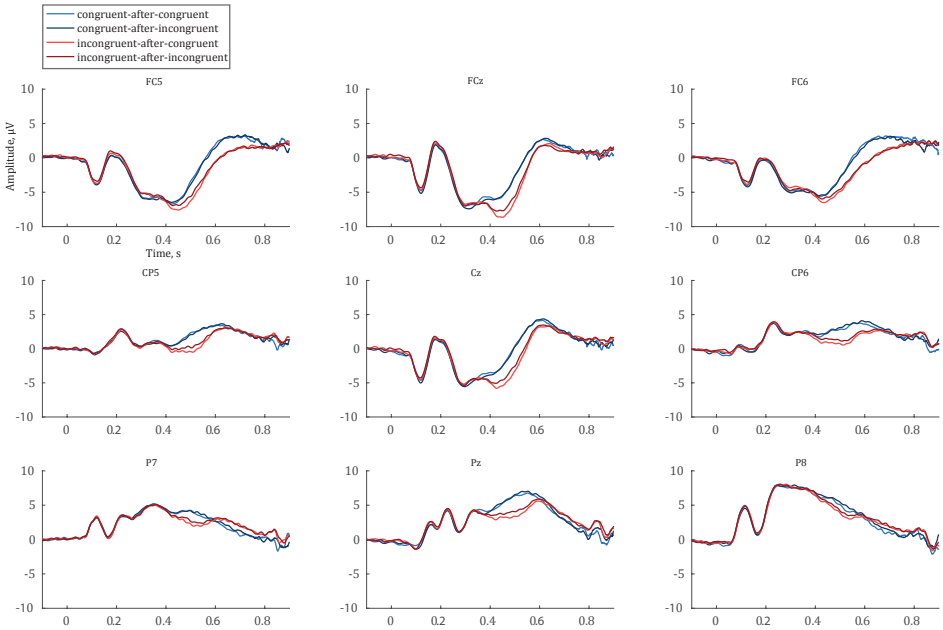

FIGURE B2. Group-averaged ERPs over nine electrode sites

APPENDIX C. SUPPLEMENTARY MATERIALS FOR CHAPTER 4

TABLE C1. Stimulus materials

Category	Picture name	Translation	Incongruent distractor	Neutral distractor
Kitchenware	pan (de)	pan	fles	vtg
	fles (de)	bottle	pan	lrpt
	bord (het)	plate	glas	vfhz
	glas (het)	glass	bord	qvjf
Clothing	trui (de)	sweater	jas	rmjx
	jas (de)	jacket	trui	dzr
	hemd (het)	vest	pak	cdlx
	pak (het)	suit	hemd	xsb
Buildings	toren (de)	tower	fabriek	rsjmx
	fabriek (de)	factory	toren	mtqbpnv
	stadion (het)	stadium	kasteel	hvftknt
	kasteel (het)	castle	stadion	bwvjzpj
Body parts	neus (de)	nose	arm	blmj
	arm (de)	arm	neus	pwc
	been (het)	leg	oog	zxlp
	oog (het)	eye	been	zls
Food	worst (de)	sausage	kaas	dtdgj
	kaas (de)	cheese	worst	xpsz
	ei (het)	egg	broodje	wd
	broodje (het)	sandwich	ei	lbzjsxs
Parts of a house	trap (de)	stairs	muur	qglj
	muur (de)	wall	trap	dhcs
	dak (het)	roof	raam	qfx
	raam (het)	window	dak	nckt
Furniture	kast (de)	cupboard	stoel	pjmh
	stoel (de)	chair	kast	dmfdr
	gordijn (het)	curtain	bureau	nzrxrrl
	bureau (het)	desk	gordijn	snpfwl

TABLE C1. Continued

Category	Picture name	Translation	Incongruent distractor	Neutral distractor
Animals	geit (de)	goat	vos	ldpc
	vos (de)	fox	geit	fkh
	hert (het)	deer	konijn	pxpr
	konijn (het)	rabbit	hert	fpcndz
Weapons	dolk (de)	dagger	speer	tqkb
	speer (de)	spear	dolk	hlptv
	kanon (het)	cannon	pistool	rbwhq
	pistool (het)	gun	kanon	brcdgtq

APPENDIX D. SUPPLEMENTARY MATERIALS FOR CHAPTER 5

TABLE D1. Stimulus materials

Category	Determiner	Noun
Animals	het	hert (deer)
	het	konijn (rabbit)
	de	geit (goat)
	de	zwaan (swan)
Kitchenware	de	fles (bottle)
	de	kan (pitcher)
	het	glas (glass)
	het	bord (plate)
Furniture	de	stoel (chair)
	de	kast (cupboard)
	het	bed (bed)
	het	bureau (desk)
Body parts	de	arm (arm)
	de	neus (nose)
	het	been (leg)
	het	oor (ear)

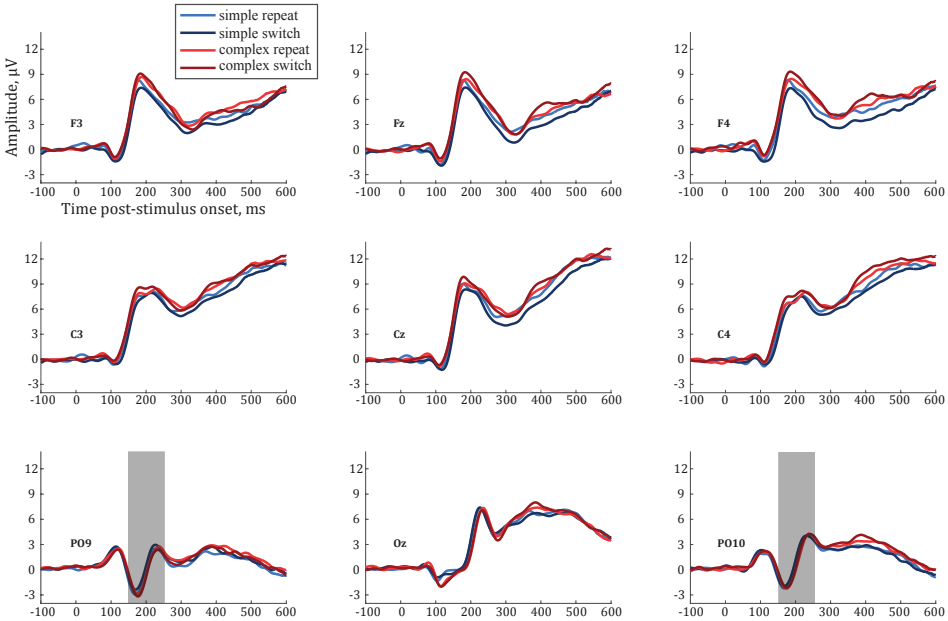

FIGURE D1. Group-averaged ERPs for nine electrode sites

REFERENCES

A

Aarts, E., & Roelofs, A. (2011). Attentional control in anterior cingulate cortex based on probabilistic cueing. *Journal of Cognitive Neuroscience, 23,* 716727.

Aarts, E., Roelofs, A., & Van Turennout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *The Journal of Neuroscience*, *28*, 4671-78.

Aristei, S., Melinger, A., & Rahman, R. A. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience*, *23*, 1567-86.

Ayora, P., Peressotti, F., Alario, F. X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: A dual-task study. *Frontiers in Psychology, 2,* 1-10.

B

Badzakova-Trajkov, G., Barnett, K. J., Waldie, K. E., & Kirk, I. J. (2009). An ERP investigation of the Stroop task: the role of the cingulate in attentional allocation and conflict resolution. *Brain research*, *1253*, 139-148.

Barbati, G., Porcaro, C., Zappasodi, F., Rossini, P. M., & Tecchio, F. (2004). Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clinical Neurophysiology*, *115*, 1220-32.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48.

Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, *123*, 84-99.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5,* 341345.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624652.

Botvinick, M. M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*, 179181.

Brooker, B. H., & Donald, M. W. (1980). Contribution of the speech musculature to apparent human EEG asymmetries prior to vocalization. *Brain and Language*, *9*, 226-245.

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). Predictive validity of the N2 and P3 ERP components to executive functioning in children: a latent-variable analysis. *Frontiers in Human Neuroscience*, *8*, 1-10.

Bürki, A., & Laganaro, M. (2014). Tracking the time course of multi-word noun phrase production with ERPs or on when (and why) cat is faster than the big cat. *Frontiers in Psychology*, *5*, 1-13.

Bürki, A., Sadat, J., Dubarry, A.-S., & Alario, F.-X. (2016). Sequential processing during noun phrase production. *Cognition*, *146*, 90-99.

C

Capizzi, M., Ambrosini, E., Arbula, S., Mazzonetto, I., & Vallesi, A. (2016). Electrophysiological evidence for domain-general processes in task-switching. *Frontiers in Human Neuroscience*, *10*, 1-14.

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 92-101.

Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., & Cohen, J. D. (2000). Parsing executive processes: Strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences USA*, *97*, 194448.

Casarotto, S., Bianchi, A. M., Cerutti, S., & Chiarenza, G. A. (2005). Dynamic time warping in the analysis of event-related potentials. *IEEE Engineering in Medicine and Biology Magazine*, *24*, 68-77.

Cavanagh, J. F., Cohen, M. X., & Allen, J. J. (2009). Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. *The Journal of Neuroscience*, *29*, 98105.

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, *49*, 220238.

Clayson, P. E., & Larson, M. J. (2011). Conflict adaptation and sequential trial effects: Support for the conflict monitoring theory. *Neuropsychologia*, *49*, 195361.

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press.

Cohen, M. X., & Cavanagh, J. F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. *Frontiers in psychology*, *2*, 1-12.

D

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345-361.

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., & Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE transactions on Biomedical Engineering*, *53*, 2583-87.

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition*, *31*, 353-359.

De Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F. X., Huffel, V. S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics*, *8*, 135-150.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *The Journal of Neuroscience*, *25*, 11730-37.

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*, 380-396.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717-722.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology*, *1*, 1-10.

Donohue, S. E., Liotti, M., Perez III, R., & Woldorff, M. G. (2012). Is conflict monitoring supramodal? Spatio-temporal dynamics of cognitive control processes in an auditory Stroop task. *Cognitive, Affective, & Behavioral Neuroscience*, *12*, 115.

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The congruency sequence effect 3.0: A critical test of conflict adaptation. *PloS one*, *9*, 18.

E

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 380390.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*, 178490.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143149.

Eulitz, C., Hauk, O., & Cohen, R. (2000). Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clinical Neurophysiology*, *111*, 2088-97.

Evans, J. L., Selinger, C., & Pollak, S. D. (2011). P3 as a measure of processing capacity in auditory and visual domains in specific language impairment. *Brain Research*, *1389*, 93-102.

F

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1058-79.

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*, 790-796.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, *45*, 152-170.

Freund, M., Gordon B., & Nozari, N. (2016). Conflict-based regulation of control in language production. *Proceedings of the 38th Annual Conference of the Cognitive Science Society,* 1625-30.

G

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 875-894.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 640-654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118*, 13-42.

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical neurophysiology*, *114*, 1580-93.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*, 480506.

H

Habets, B., Jansma, B. M., & Münte, T. F. (2008). Neurophysiological correlates of linearization in language production. *BMC Neuroscience*, *9*, 1-8.

Hanslmayr, S., Pastötter, B., Bäuml, K. H., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience, 20*, 215225.

Hashimoto, N., & Thompson, C. K. (2010). The use of the picture–word interference paradigm to examine naming abilities in aphasic individuals. *Aphasiology*, *24*, 580-611.

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and cognitive processes*, 8, 379-411.

Hommel, B., Proctor, R. W., & Vu, K. P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*, 117.

Huang, H. C., & Jansen, B. H. (1985). EEG waveform analysis by means of dynamic time-warping. *International journal of bio-medical computing*, *17*, 135-144.

Hung, C. L., Huang, C. J., Tsai, Y. J., Chang, Y. K., & Hung, T. M. (2016). Neuroelectric and behavioral effects of acute exercise on task switching in children with attention-deficit/hyperactivity disorder. *Frontiers in Psychology*, *7*, 1-11.

I

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology*, *2,* 1-16.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101-144.

J

Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition*, *4*, 169-178.

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 249-256.

K

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall.

Kamijo, K., & Takeda, Y. (2010). Regular physical activity improves executive function during task switching in young adults. *International Journal of Psychophysiology*, *75*, 304-311.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 102326.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language*, *47*, 145-171.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557-577.

Kopp, B., Rist, F., & Mattler, U. W. E. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*, 282294.

Kramer, A. F., Sirevaag, E. J., & Braune, R. (1987). A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors*, *29*, 145-160.

Kutas M., Van Petten C., & Kluender R. Psycholinguistics electrified II. In: Traxler, M., & Gernsbacher, M. A. (Eds.). (2011). *Handbook of psycholinguistics.* Academic Press.

L

Lamers, M. J., & Roelofs, A. (2011). Attentional control adjustments in Eriksen and Stroop task performance can be independent of response conflict. *The Quarterly Journal of Experimental Psychology*, *64*, 105681.

Larson, M. J., Clawson, A., Clayson, P. E., & South, M. (2012). Cognitive control and conflict adaptation similarities in children and adults. *Developmental neuropsychology*, *37*, 343357.

Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and critique of conflict-related ERPs. *International Journal of Psychophysiology*, *93*, 283297.

Larson, M. J., Clayson, P. E., Kirwan, C. B., & Weissman, D. H. (2016). Event-related potential indices of congruency sequence effects without feature integration or contingency learning confounds. *Psychophysiology*, *53*, 814822.

Larson, M. J., Kaufman, D. A., & Perlstein, W. M. (2009). Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia*, *47*, 663670.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*, 1-38.

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, *38*, 701-711.

Lorist, M.M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., & Meijman, T.F. (2000). Mental fatigue and task control: Planning and preparation. *Psychophysiology*, *37*, 614-625.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291-308.

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, *115*, 1821-35.

M

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, *109*, 163-203.

Marek, A., Habets, B., Jansma, B. M., Nager, W., & Münte, T. F. (2007). Neural correlates of conceptualisation difficulty during the preparation of complex utterances. *Aphasiology*, *21*, 1147-56.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164,* 177-190.

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *Journal of Neuroscience*, *28*, 12539-45.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*, 450452.

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*, 477-496.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological review*, *104*, 3-65.

Michel Lange, V. M., Perret, C., & Laganaro, M. (2015). Comparison of single-word and adjective-noun phrase production using event-related brain potentials. *Cortex*, *67*, 15-29.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49-100.

N

Nigbur, R., Ivanova, G., & Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical Neurophysiology*, *122*, 218594.

Notebaert, W., Gevers, W., Verbruggen, F., & Liefooghe, B. (2006). Top-down and bottom-up sequential modulations of congruency effects. *Psychonomic Bulletin & Review*, *13*, 112117.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*, 1-33.

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science, 26,* 403-410.

O

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*, 1-9.

Ouyang, G., Sommer, W., & Zhou, C. (2015a). A toolbox for residue iteration decomposition (RIDE)—a method for the decomposition, reconstruction, and single trial analysis of event related potentials. *Journal of neuroscience methods*, *250*, 7-21.

Ouyang, G., Sommer, W., & Zhou, C. (2015b). Updating and validating a new framework for restoring and analyzing latency-variable ERP components from single trials with residue iteration decomposition (RIDE). *Psychophysiology*, *52*, 839-856.

Ouyang, G., Sommer, W., Zhou, C., Aristei, S., Pinkpank, T., & Rahman, R. A. (2016). Articulation artifacts during overt language production in event-related brain potentials: Description and correction. *Brain topography*, *29*, 791-813.

P

Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 358-377.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, *116*, 220-244.

Pastötter, B., Dreisbach, G., & Bäuml, K. H. T. (2013). Dynamic adjustments of cognitive control: oscillatory correlates of the conflict adaptation effect. *Journal of Cognitive Neuroscience*, *25*, 2167-78.

Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, *64*, 404-423.

Piai, V., Roelofs, A., & Schriefers, H. (2012). Distractor strength and selective attention in picture-naming performance. *Memory & Cognition*, *40*, 614-627.

Piai, V., Roelofs, A., & Schriefers, H. (2014). Locus of semantic interference in picture naming: Evidence from dual-task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 147-165.

Piai, V., Roelofs, A., & van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain research, 1450,* 87-101.

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J. M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLoS One, 9,* 1–11

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neuropsychology, 118,* 2128-48.

Polich, J., Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology, 41,* 103-146.

Porcaro, C., Medaglia, M. T., & Krott, A. (2015). Removing speech artifacts from electroencephalographic recordings during overt picture naming. *NeuroImage, 105,* 171-180.

Proverbio, A. M., Burco, F., del Zotto, M., & Zani, A. (2004). Blue piglets? Electrophysiological evidence for the primacy of shape over color in object recognition. *Cognitive Brain Research, 18,* 288-300.

Pylkkänen, L., Bemis, D. K., & Elorrieta, E. B. (2014). Building phrases in language production: An MEG study of simple composition. *Cognition, 133,* 371-384.

R

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review, 107,* 460-499.

Rayner, K., & Springer, C. J. (1986). Graphemic and semantic similarity effects in the picture-word interference task. *British Journal of Psychology, 77,* 207-222.

Riès, S., Janssen, N., Burle, B., & Alario, F. X. (2013). Response-locked brain dynamics of word production. *PLoS One, 8,* 1-14

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107-142.

Roelofs, A. (2000). WEAVER++ and other computational models of lemma retrieval and word-form encoding. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 71-114). Sussex, UK: Psychology Press.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological review, 110,* 88-125.

Roelofs, A. (2014). Tracking eye movements to localize Stroop interference in naming: Word planning versus articulatory buffering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1332-47.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2,* 1-14.

Roelofs, A., & Shitova, N. (2017). Importance of response time in assessing the cerebral dynamics of spoken word production: Comment on Munding et al. (2016). *Language, Cognition and Neuroscience, 32,* 1064-69.

Roelofs, A., Piai, V., & Schriefers, H. (2013). Context effects and selective attention in picture naming and word reading: Competition versus response exclusion. *Language and Cognitive Processes, 28*, 655-671.

Roelofs, A., Piai, V., Garrido Rodriguez, G., & Chwilla, D. J. (2016). Electrophysiology of cross-language interference and facilitation in picture naming. *Cortex, 76*, 116.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207-231.

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development, 46,* 247-253.

S

Schnur, T. T., & Martin, R. (2012). Semantic picture–word interference is a postperceptual effect. *Psychonomic Bulletin & Review, 19*, 301-308.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition, 45*, 33-54.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language, 29*, 86-102.

Seiger-Gardner, L., & Schwartz, R. G. (2008). Lexical access in children with and without specific language impairment: a cross-modal picture–word interference study. *International Journal of Language & Communication Disorders, 43,* 528-551.

Shitova, N., Roelofs, A., Coughler, C., & Schriefers, H. (2017). P3 event-related brain potential reflects allocation and use of central processing capacity in language production. *Neuropsychologia, 106*, 138-145.

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M. (2016). Using brain potentials to functionally localise Stroop-like effects in colour and picture naming: Perceptual encoding versus word planning. *PloS one, 11*, 1-16.

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M. (2017). Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex, 92*, 289-303.

Sikora, K., Roelofs, A., & Hermans, D. (2016). Electrophysiology of executive control in spoken noun-phrase production: Dynamics of updating, inhibiting, and shifting. *Neuropsychologia, 84*, 44-53.

Sikora, K., Roelofs, A., Hermans, D., & Knoors, H. (2016). Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. *The Quarterly Journal of Experimental Psychology, 69*, 1719-40.

Strayer, D. L., & Drews, F. A. (2007). Cell-phone–induced driver distraction. *Current Directions in Psychological Science, 16*, 128-131.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20*, 912-928.

Strijkers, K., Holcomb, P.J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language, 65*, 345-362.

Swainson, R., Cunnington, R., Jackson, G. M., Rorden, C., Peters, A. M., Morris, P. G., & Jackson, S. R. (2003). Cognitive control mechanisms revealed by ERP and fMRI: Evidence from repeated task-switching. *Journal of Cognitive Neuroscience, 15*, 785-799.

T

Thompson, C. K., Cho, S., Price, C., Wieneke, C., Bonakdarpour, B., Rogalski, E., ... & Mesulam, M. M. (2012). Semantic interference during object naming in agrammatic and logopenic primary progressive aphasia (PPA). *Brain and language*, *120*, 237-250.

U

Ueno, T., Saito, S., Rogers, T. T., & Ralph, M. A. L. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*, 385-396.

Ullsperger, M., Bylsma, L. M., & Botvinick, M. M. (2005). The conflict adaptation effect: It's not just priming. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 467472.

V

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudo-randomization. *Behavior research methods*, *38*, 584-589.

Van Maanen, L., & Van Rijn, H. (2010). The locus of the Gratton effect in picture–word interference. *Topics in Cognitive Science, 2*, 168180.

Van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review*, *16*, 987-999.

Van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive science*, *36*, 62-101.

Van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: a functional MRI study. *Neuroimage*, *27*, 497-504.

Verguts, T., Notebaert, W., Kunde, W., & Wühr, P. (2011). Post-conflict slowing: Cognitive adaptation after conflict processing. *Psychonomic Bulletin & Review*, *18*, 76-82.

W

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P3 morphology under divided attention. *Psychophysiology*, *38*, 998-1003.

West, R., & Bailey, K. (2012). ERP correlates of dual mechanisms of control in the counting Stroop task. *Psychophysiology*, *49*, 130918.

West, R., Bailey, K., Tiernan, B. N., Boonsuk, W., & Gilbert, S. (2012). The temporal dynamics of medial and lateral frontal neural activity related to proactive cognitive control. *Neuropsychologia*, *50*, 345060.

X

Xiao, X., Zhang, Q., Jia, L., Zhang, Y., & Luo, J. (2010). Temporal course of cognitive control in a picture–word interference task. *Neuroreport*, *21*, 104-107.

Y

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931-959.

NEDERLANDSE SAMENVATTING

Het benoemen van een voorwerp in een afbeelding vereist dat de spreker of spreekster van een taal tot een aantal zaken in staat is: ten eerste moet hij of zij begrijpen welk voorwerp afgebeeld wordt, vervolgens moet hij of zij een woord vinden om het voorwerp accuraat te omschrijven, en tenslotte moet hij of zij het woord uitspreken. Het proces van de keuze van een woord voor een voorwerp heet lexicale selectie. Hoewel het voor de meeste gezonde volwassenen moeiteloos verloopt, is lexicale selectie een complex mechanisme dat snel verstoord kan raken, bijvoorbeeld na een beroerte. De reden hiervoor is dat woorden niet op zichzelf staan in onze lexicons, maar verbonden zijn in netwerken, gebaseerd op de mate van overeenkomst tussen de voorwerpen die zij vertegenwoordigen. De woorden *hond* en *kat* staan bijvoorbeeld dicht bij elkaar, *vis* staat een beetje verder van weg, en *raam* staat heel ver weg. Indien een spreker *hond* wil uitspreken, wordt ook het woord *kat* geactiveerd in hun lexicale netwerk (en in mindere mate ook *vis*). Zodoende is *hond* dus in competitie met *kat* om geselecteerd te worden.

Deze competitie kan nader onderzocht worden in een experimentele opzet. De opdracht om een voorwerp te benoemen kan aangepast worden om de lexicale selectie moeilijker te maken. Het zogenaamde plaatje-woord interferentie paradigma (*picture-word interference paradigm, PWI*) maakt gebruik van gesuperponeerde afleidende woorden. Bij een afleidend woord dat *congruent* is met het voorwerp in de afbeelding (bijvoorbeeld het woord *hond* in de afbeelding van een hond), is het benoemen makkelijker dan bij een afleidend woord dat *incongruent* is met het voorwerp (bijvoorbeeld het woord *kat* gesuperponeerd op een afbeelding van een hond). Dit komt doordat bij congruente stimuli de afbeelding en het afleidende woord hetzelfde correcte woord activeren, terwijl bij incongruente stimuli enkel de afbeelding het correcte woord activeert (*hond*), terwijl het afleidende woord een competitie-woord activeert (*kat*). Op deze wijze is de PWI opdracht vergelijkbaar met de beroemde Stroop opdracht: het benoemen van de kleur waarin een woord staat afgedrukt is in het algemeen eenvoudiger wanneer dat woord overeenkomt met de kleur die benoemd wordt (bijvoorbeeld ***blauw***), dan wanneer dit niet het geval is (bijvoorbeeld ***rood***). Onderzoek heeft aangetoond dat de effecten van het afleidend woord in de PWI en de Stroop taken door soortgelijke mechanismen in de lexicale competitie tot stand komen.

Deze overeenkomst tussen de PWI en de Stroop test lijkt intuïtief, maar werd door recente experimentele gegevens betwist. Deze suggereerden dat, in tegenstelling tot de Stroop test, het PWI effect niet verbonden is met competitieve lexicale selectie, maar met een vroegere fase van stimulusverwerking, wanneer voorwerpen worden waargenomen en vergeleken met bestaande concepten in het langetermijngeheugen. Hoofdstuk 2 van dit proefschrift beschrijft een onderzoek dat direct gericht is op de vraag of vergelijkbare competitieve

mechanismen in de fase van lexicale selectie bestaan in de PWI en de Stroop opdrachten. De Stroop-achtige effecten in de Stroop en PWI opdracht werden direct vergeleken in dezelfde groep onderzoekspersonen en met gelijke experimentele parameters. Wij gebruikten de *Event-related Potentials* techniek (ERP, gebaseerd op EEG opnamen) om de timing van het Stroop-achtige effect tussen opdrachten te vergelijken. De Stroop-achtige effecten in zowel Stroop als PWI waren geassocieerd met modulatie binnen het tijdsvenster van lexicale selectie, ongeveer 350-500 milliseconden na het vertonen van de afbeelding, terwijl geen vroegere effecten zichtbaar waren. Deze resultaten suggereren dat het Stroop-achtige effect is verbonden met de fase van lexicale selectie in de verwerkingsstroom, onafhankelijk van het type stimulus (afbeelding of kleur). Dit patroon van resultaten biedt nieuwe ondersteuning voor de competitieve verklaring.

Het Stroop-effect kan groter of kleiner zijn afhankelijk van het vermogen en de bereidheid van de spreker om de overheersende respons ingegeven door het afleidende woord te onderdrukken. Indien de spreker goede controle heeft over zijn antwoorden, is het Stroop-achtige effect kleiner. Opmerkelijkerwijs kan de mate van controle binnen een spreker veranderen: indien een spreker een makkelijke (congruente) stimulus verwacht, oefent hij minder controle uit dan bij een moeilijke (incongruente) stimulus. Verder suggereert een experimentele bevinding in de literatuur dat afname van controle het Stroop-achtig effect verplaatst van het lexicale selectiestadium naar het perceptiestadium. Deze bevinding staat op gespannen voet met de competitie verklaring. In Hoofdstuk 3 presenteerde ik een PWI studie waar het afleidende effect werd onderzocht als een functie van het vorige type stimulus. De verwachting was dat het Stroop-achtige effect kleiner zou zijn indien de vorige stimulus incongruent was dan wanneer deze congruent was. In lijn met het vorige onderzoek, beschreven in Hoofdstuk 2, werd het afleidend effect waargenomen tijdens de lexicale selectie, hetgeen zowel voor post-congruente (weinig controle) en post-incongruente stimuli (veel controle) gold. Door modulatie van controle was het Stroop-achtige effect echter kleiner bij post-incongruente stimuli dan bij post-congruente stimuli, precies als verwacht. Derhalve bevestigde het onderzoek gepresenteerd in Hoofdstuk 3 dat de competitieve verklaring onafhankelijk van een ervaren conflict in de vorige stimulus van kracht blijft.

De aanpassingen in controle bij het benoemen van een afbeelding werden verder onderzocht in Hoofdstuk 4, waar ik mij richtte op de aard van de controle aanpassingen bij het spreken. Er bestaan twee verklaringen voor dit fenomeen. Een dominante verklaring (*conflict monitoring*) stelt dat afname van het Stroop-achtige effect op post-incongruente stimuli, vergeleken met post-congruente stimuli, wordt veroorzaakt door het ervaren conflict in voorafgaande incongruente stimuli. Een dergelijke versterkte controle stelt de spreker vermoedelijk beter in staat om de afleidende informatie in de volgende stimulus te onderdrukken. Een alternatieve verklaring (*congruency expectancy*) suggereert dat zo'n gedragspatroon eerder voortvloeit uit

de verwachting van congruentie in post-congruente stimuli: indien het afleidende woord in de vorige afbeelding in werkelijkheid hielp bij het benoemen van het voorwerp, zal de spreker de afleidende informatie sterker "vertrouwen" in de volgende stimulus, waardoor hij sneller wordt, als de volgende stimulus ook daadwerkelijk congruent is. In een serie van drie gedragsexperimenten onderzochten wij controle aanpassingen bij enkel-woord en zinsdeelproductie tijdens de PWI opdracht. Op basis van waargenomen gedragspatronen in de resultaten konden wij concluderen dat de aanpassing van controle in de PWI eerder het gevolg is van verwachte congruentie dan post-conflict adaptatie in de verwerking.

Aanpassingen in top-down controle afhankelijk van de verwachting dat een congruente of incongruente PWI stimulus verwacht wordt, is slecht één van vele situaties waarin een spreker of spreekster anticipeert op een aankomende taak en zijn of haar gedrag strategisch aanpast om goed te presteren. Meer algemeen gesproken moet een spreker of spreekster een zekere mate van capaciteit steken in het spreken afhankelijk van de moeilijkheidsgraad van de opdracht. Als de situatie vereist dat de spreker of spreekster een voorwerp uitgebreid omschrijft ("de grote rode beker"), zal dit meer verwerkingscapaciteit kosten dan wanneer een simpele verwijzing voldoende is om het idee te verwoorden ("de beker"). In Hoofdstuk 5 beschrijf ik electrofysiologische patronen van toekenning en gebruik van verwerkingscapaciteit in zinsdeelproductie. Het gepresenteerd onderzoek toonde aan dat sprekers de toekenning van verwerkingscapaciteit kunnen aanpassen op basis van de beschikbare informatie over de complexiteit van een aankomende opdracht. Daarbij komen zulke aanpassingen in de hoeveelheid toegekende verwerkingscapaciteit tot uitdrukking in dezelfde ERP component die gebruik van verwerkingscapaciteit toont.

Concluderend heb ik in een serie van studies de competitieve verklaring in enkel-woord en zinsdeelproductie opdrachten getest. Meer in het bijzonder heb ik evidentie gepresenteerd ten gunste van het model van spraakproductie dat beweert dat het afleidend effect in de Stroop-achtige opdrachten, inclusief PWI, is verbonden met het stadium waarin het woord worden geselecteerd. Verder heb ik onderzocht hoe de verwachtingen van onderzoekspersonen aangaande het aankomende type stimulus hun strategieën voor de opdracht beïnvloedt. Ik presenteerde onderzoek naar aanpassingen in controle van woordplanning in de PWI opdracht en naar aanpassingen in de toekenning van capaciteit in een zinsdeelproductie opdracht. Onderzoek naar de aard van controle aanpassingen in PWI, gepresenteerd in dit proefschrift, tonen aan dat zulke aanpassingen het gevolg zijn van verwachting van congruentie en primair worden gedreven door congruente stimuli, niet door incongruente stimuli, zoals gesuggereerd door de dominante conflict-monitoring verklaring. Daarnaast suggereerde analyse van hersenactiviteitspatronen geassocieerd met controle aanpassing in PWI dat Stroop-achtige effecten verbonden zijn met lexicale selectie, hetgeen nieuwe evidentie levert voor de competitieve verklaring. Tenslotte besprak ik in de laatste studie

een ander type aanpassingen gedreven door verwachting, namelijk capaciteitstoekenning. In tegenstelling tot de hiervoor beschreven situatie in de PWI opdracht waren de proefpersonen zich bewust van het type van de aankomende stimulus en konden zij hun verwachtingen bijstellen om een meer of minder complexe afbeelding te verwerken, en tevens een meer of minder uitgebreid antwoord te formuleren. De resultaten toonden dat het vermogen om zich aan te passen aan de complexiteit van de aankomende stimulus de prestatie van de proefpersonen zowel in de fase van perceptie (de vroege fase van woordverwerking) als in de fase van linguïstische verwerking beïnvloedde.

ACKNOWLEDGEMENTS

Dear Ardi, thank you for your guidance and inspiration through all these four years. You were always thinking along and you have always been enthusiastic, encouraging, and positive.

Dear Herbert, I am so grateful to you for teaching me to see my work on all possible scales and from all possible angles. Your experience helped me treat ups and downs philosophically.

Dear Jan-Mathijs, thank you for teaching me to love data as much as you do. You showed me that "not working" just does not exist, only "not working yet". You made me feel empowered.

Dear Marcel, I appreciate your valuable advice no matter how improbable my question was.

Dear IMPRS community, Dirkje, Els, and Kevin, thank you for brightening my PhD years with a lot of great activities, both educational and social.

Dear TSG, Pascal, Wilbert, and Gerard, you were always ready to help and fix the impossible. Thank you!

Dear Miriam, I am so grateful to you for sharing your knowledge and experience with me, it was a great pleasure to work with you. I am very happy you supported me especially when it came to tough career-changing decisions.

Dear Vitória, thank you for always inspiring me to aim higher. You asked intelligent questions and you gave honest advice. I really appreciate your kindness and support.

Dear Sara, I loved our deep, encouraging, and sincere conversations. Thank you for always believing in me.

Dear Fenny, I am so grateful to you for teaching me to find time for great things in life and to value people, art, and nature the way you do. I hope we will stay in touch no matter where on this planet you choose to be.

Dear Sybrine, thank you for sharing with me all your joy, energy, and love for everything that is new and lively.

Dear Lily, you always had words of support for me, and the emotional sharing that we did made me grow as a person. Thank you!

Dear Annika, you were a great officemate for me. I appreciate what a good listener you were (like you had a choice), and how you reviewed all my posters and slides, and, not to forget, how you taught me all these useless yet spectacular German words!

Dear Natalia, thank you for sharing your ideas so generously and for being so honest in everything you say or do.

Dear Indah, morning meditation was probably the most important discovery for me in the last four years. I am very grateful to you for your guidance, advice, appreciation, and attention. Dear Mareike, Sara, Julia, Francesca, thank you for meditating along, sharing and supporting.

Dear Donders Volley team - Alex, Asustosh, Claudia, Dan, Femke, Fenny, Jil, Kristjan, Larry, Linda, Loes, Marvin, Matthias, Pim, Ruud, Sanne, Sara, Sasha, Sophie, Steffie, Sybrine, Tobi, Xiaochen - thank you for all this amazing fun!

Dear Language division friends - Anne, Annika, Arushi, Jana, Johanna, Kasia, Kevin, Marpessa, Monica, Nikki, Sybrine, Wendy, Xiaochen - it was inspiring to talk with you in meetings, or during lunches, or while cycling home. Whether sharing a little laugh or a big frustration, we were a team. Dear Andrea, Casper, Ileana, Jana, Josi, Lara, Remco, Ricarda, Stefania, Suhas, Syannah, thank you for all the happy social moments we shared.

Dear Samarth, I never had a friend like you. Thank you for the greatest talks, the toughest arguments, the best shopping, the longest *Friends* marathons, and the tastiest pasta aglio e olio ever. You are marvellous.

Dear Hugo and Monika, and of course, little Emma, thank you for the great time that we had together in the Netherlands and in Germany and for all the interesting and inspiring conversations!

Dear Irisha and Petya, and of course, little Sasha, thank you for all our great trips, board game evenings, and a wonderful family time in Zurich. I really appreciate your support and encouragement.

Lieve Diny en Jan, jullie zijn de eerste mensen die mij in Nederland thuis lieten voelen. Bedankt voor al jullie vriendelijkheid, respect, en aandacht. Jullie ondersteuning was heel belangrijk voor mij. En nogmaals dank voor het ontwerpen van de omslag van mijn proefschrift, Jan!

Дорогие мама, папа и бабушка, без вас я бы не была там, где я сейчас, не была бы тем, кто я сейчас, не говоря уже о том, что я бы вообще не была. Я знаю, что многие мои решения было непросто принять, но вы отлично справились. Спасибо, что вы всегда на моей стороне! Люблю вас бесконечно.

Dear Rutger, thank you for unbelievable amount of support, encouragement, and love that you gave me over the years. You reminded me to breathe on those particularly stressful days. You were so confident and optimistic about my career and life, that I had no other choice than to start being confident and optimistic myself. You make me ridiculously happy.

BIOGRAPHY

Natalia Shitova was born in 1988 in Arzamas-16, USSR (now Sarov, Russia), where she graduated from high school in 2006. She continued her education at the Faculty of Philology, Lomonosov Moscow State University, and she completed her Masters in Theoretical and Computational Linguistics in 2011 (cum laude). She started a job as a computational linguist and software developer long before graduation and pursued this career until 2012, when she moved to San Sebastian (Spain) to perform Masters in Cognitive Neuroscience of Language at the Basque Center on Cognition, Brain and Language, which she completed in 2013. Same year she started her PhD project at the Donders Institute for Brain, Cognition and Behaviour at the Radboud University Nijmegen. In her PhD project Natalia investigated lexical selection in language production using behavioral and electrophysiological measures. Currently, Natalia is working as a software developer in search engine optimization at Yoast (Wijchen, The Netherlands).

PUBLICATIONS

Shitova, N. & Roelofs, A. (under revision). Speaking under distraction: Control adjustments in noun phrase production.

Shitova, N., Roelofs, A., Coughler, C., & Schriefers, H. (2017). P3 event-related brain potential reflects allocation and use of central processing capacity in language production. *Neuropsychologia*, *106*, 138-145.

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M. (2017). Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex*, *92*, 289-303.

Roelofs, A., & Shitova, N. (2017). Importance of response time in assessing the cerebral dynamics of spoken word production: Comment on Munding et al. (2016). *Language, Cognition and Neuroscience, 32,* 1064-69.

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J. M. (2016). Using brain potentials to functionally localise Stroop-like effects in colour and picture naming: Perceptual encoding versus word planning. *PloS one*, *11*, 1-16.

DONDERS GRADUATE SCHOOL FOR COGNITIVE NEUROSCIENCE SERIES

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc. Positions outside academia spread among the following sectors: - specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology, - specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy, - higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:

http://www.ru.nl/donders/graduate-school/phd/

DONDERS
INSTITUTE

Max Planck Institute
for Psycholinguistics

ISBN 978-94-6284-139-0

Radboud University  Radboudumc